

Statistics for High Energy Physics

1. Profile Likelihood, Asimov, CLs & Exclusion
2. Discovery and the Look Elsewhere Effect

Eilam Gross,
Weizmann Institute of Science



Your Lecturer



Eilam Gross

*Professor of Physics
Department of Particle Physics and Astrophysics
Weizmann Institute of Physics*
Eilam.Work@gmail.com
Eilam.gross@weizmann.ac.il

My research: Applications of DL in HEP

Search for (yet undiscovered) Higgs decaying to Charm Quarks and tagging heavy quark flavors (in particular Charm).

- 80s-90s: CERN, LEP
OPAL Higgs Convener
- 2000s : TESLA
LC WS Higgs Convener
- 2010s. : ATLAS, LHC
Statistics Convener
Higgs Convener (2012)
LHC Higgs
Combination Convener
- Present. :
Charm Physics and ML



Basically ITS ALL ABOUT NUMBERS

$$n_{\text{expected}} = L \cdot \sigma \cdot \text{eff}$$

$$[L] = \text{events/cm}^2$$

$$[\sigma] = \text{cm}^2$$



Preliminaries



A counting experiment

- The Higgs hypothesis is that of signal $s(m_H)$

$$s(m_H) = L\sigma_{SM} \cdot \epsilon$$

For simplicity unless otherwise noted

$$s(m_H) = L\sigma_{SM}$$

- In a counting experiment

$$n = \mu s(m_H) + b$$

$$\mu = \frac{L\sigma_{obs}(m_H)}{L\sigma_{SM}(m_H)} = \frac{\sigma_{obs}(m_H)}{\sigma_{SM}(m_H)}$$

- μ is the strength of the signal
(with respect to the expected Standard Model one)
- The hypotheses are therefore denoted by H_μ
- H_1 is the SM with a Higgs, H_0 is the background only model



A Tale of Two Hypotheses

NULL

ALTERNATE

- Test the Null hypothesis and try to reject it
- Fail to reject it OR reject it in favour of the alternative hypothesis



A Tale of Two Hypotheses

NULL

H_0 - SM w/o Higgs

ALTERNATE

H_1 - SM with Higgs

- Test the Null hypothesis and try to reject it
- Fail to reject it OR reject it in favour of the alternative hypothesis



A Tale of Two Hypotheses

NULL

H_0 - SM w/o Higgs

ALTERNATE

H_1 - SM with Higgs

Rejecting H_0 in favour of $H_1(m_H) \rightarrow$ Discovery of a Higgs with a mass m_H

We quantify rejection by p-value (later)



Swapping Hypotheses \rightarrow exclusion

NULL

H_0 - SM w/o Higgs

ALTERNATE

H_1 - SM with Higgs

- Reject H_1 in favor of H_0

Excluding H_1 (m_H) \rightarrow Excluding the Higgs with a mass m_H

We quantify rejection by p-value (later)



Likelihood

- Likelihood is the compatibility of the Hypothesis with a given data set.
But it depends on the data

Likelihood is not the probability of the hypothesis given the data

$$L(H) = P(x | H)$$

$$P(x | H) \neq P(H | x)$$

Bayes Theorem

$$P(H | x) = \frac{P(x | H) \cdot P(H)}{\sum_H P(x | H) P(H)}$$

$$P(H | x) \approx P(x | H) \cdot P(H)$$

Prior



Testing an Hypothesis (wikipedia...)

- The first step in any hypothesis test is to state the relevant **null**, H_0 and **alternative hypotheses**, say, H_1
- The next step is to define a test statistic, q , **under the null hypothesis**
- Compute from the observations the observed value q_{obs} *of the test statistic* q .
- Decide (based on q_{obs}) to **either**
fail to reject the null hypothesis **or**
reject it in favor of an alternative hypothesis
- **next: How to construct a test statistic, how to decide?**



Test statistic and p-value



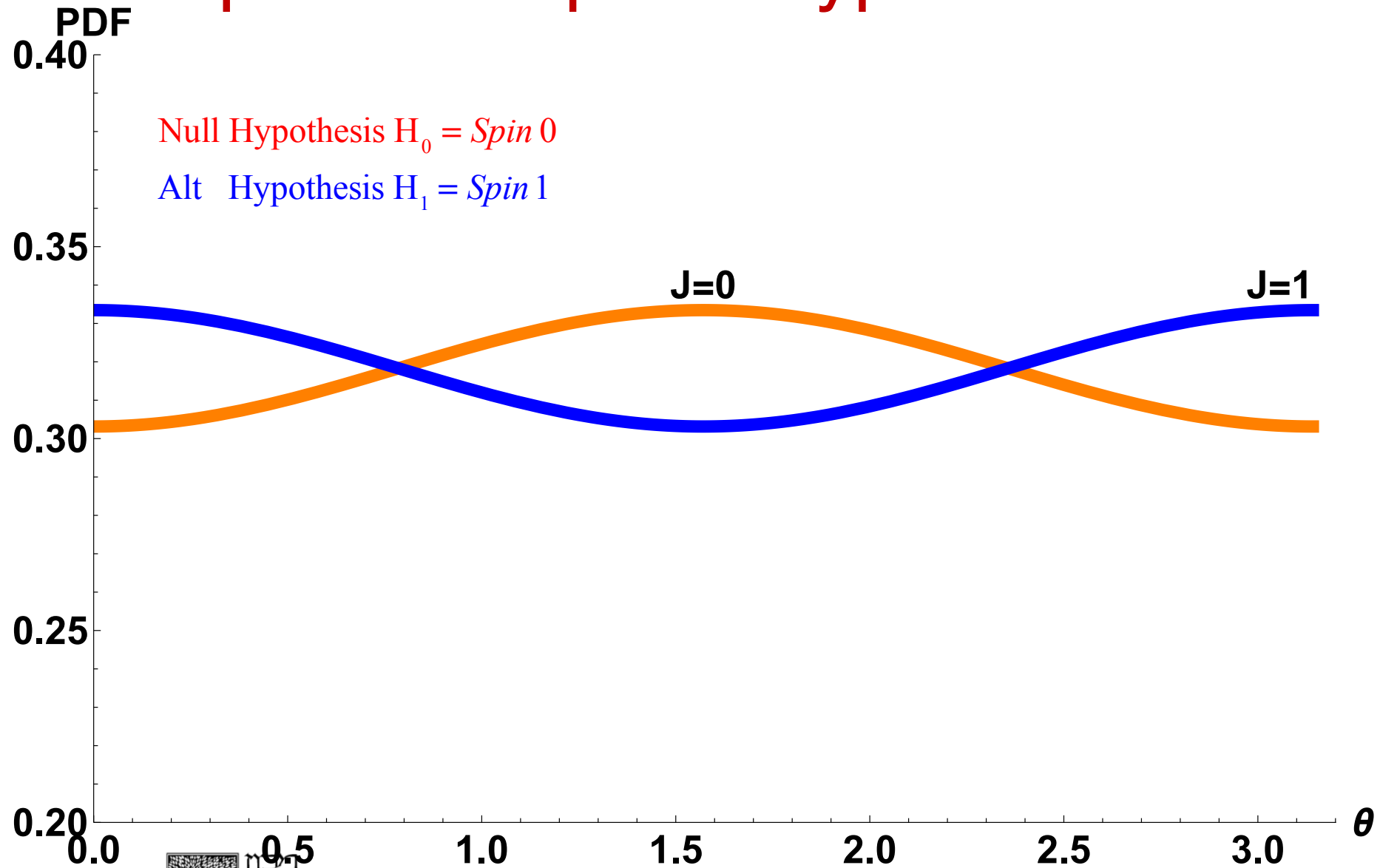
Case Study 1 : Spin



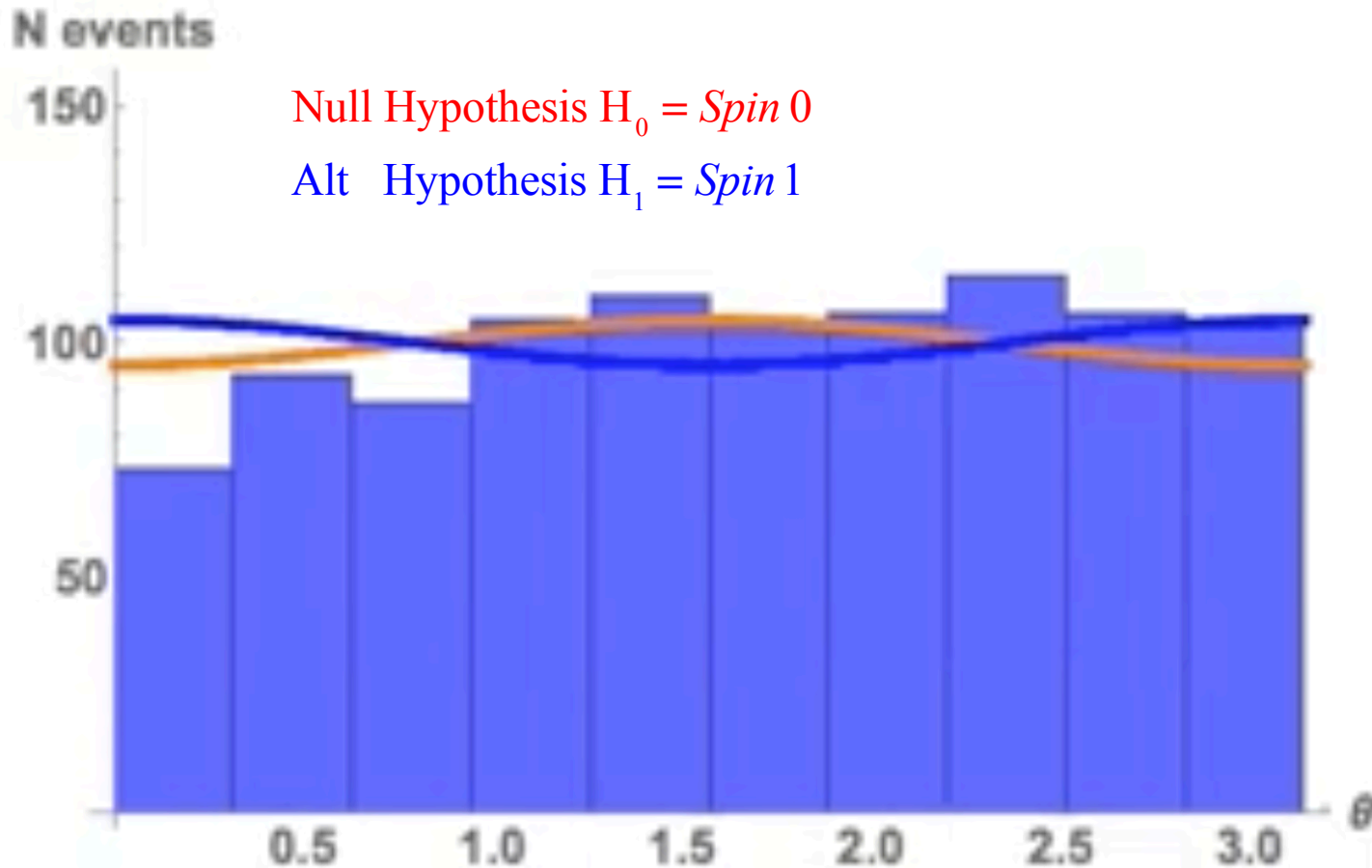
Spin 0 vs Spin 1 Hypotheses

Null Hypothesis $H_0 = \text{Spin } 0$

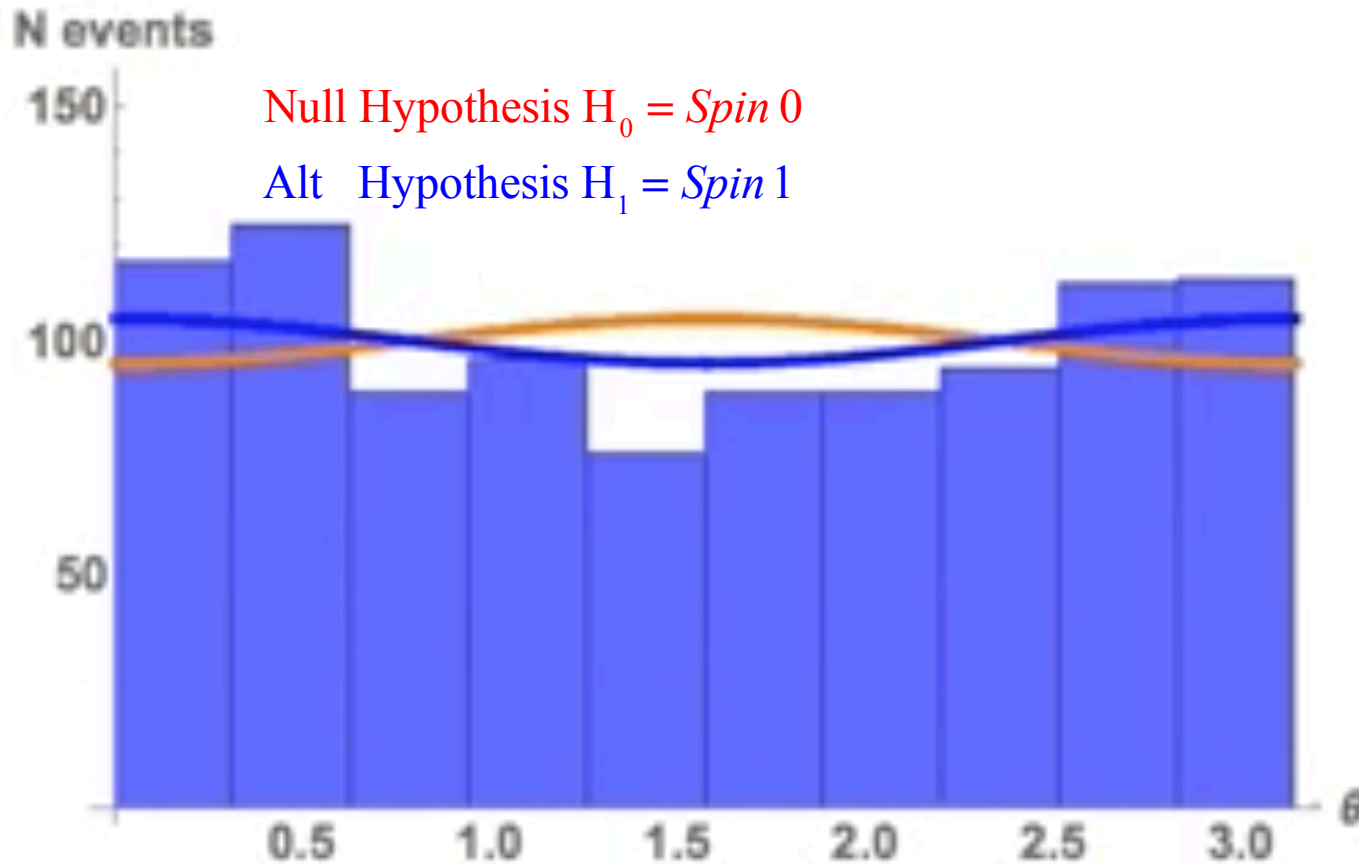
Alt Hypothesis $H_1 = \text{Spin } 1$



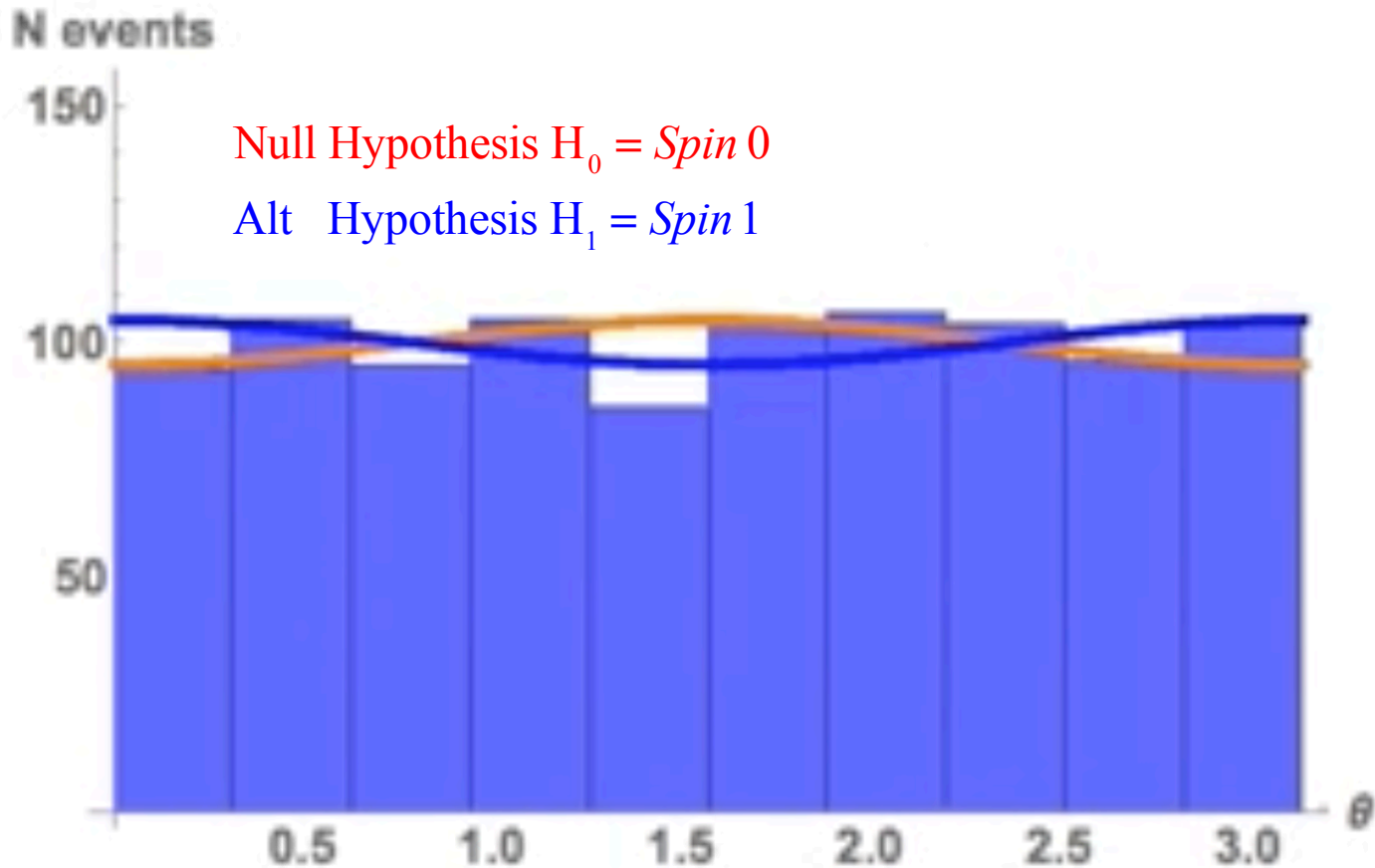
Spin 0 vs Spin 1 Hypotheses



Spin 0 vs Spin 1 Hypotheses



Spin 0 vs Spin 1 Hypotheses



The Neyman-Pearson Lemma

- Define a **test statistic** $\lambda = \frac{L(H_1)}{L(H_0)}$
- When performing a hypothesis test between two simple hypotheses, H_0 and H_1 ,
the Likelihood Ratio test, $\lambda = \frac{L(H_1)}{L(H_0)}$

which rejects H_0 in favor of H_1 ,

is the most powerful test

for a given significance level
with a threshold η

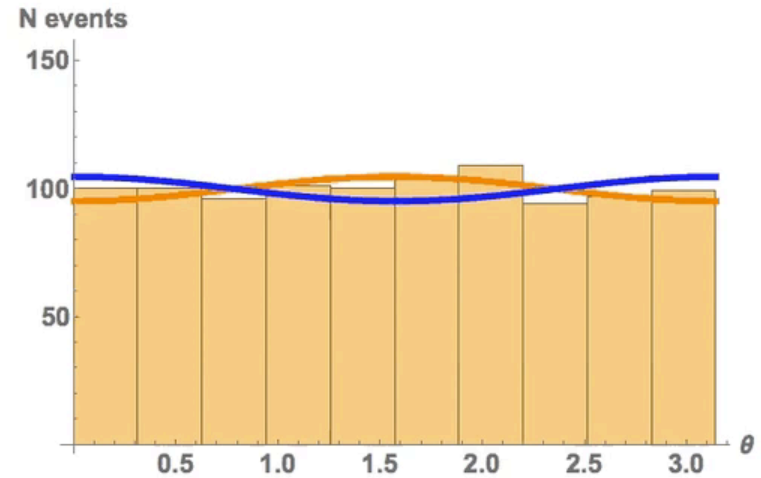
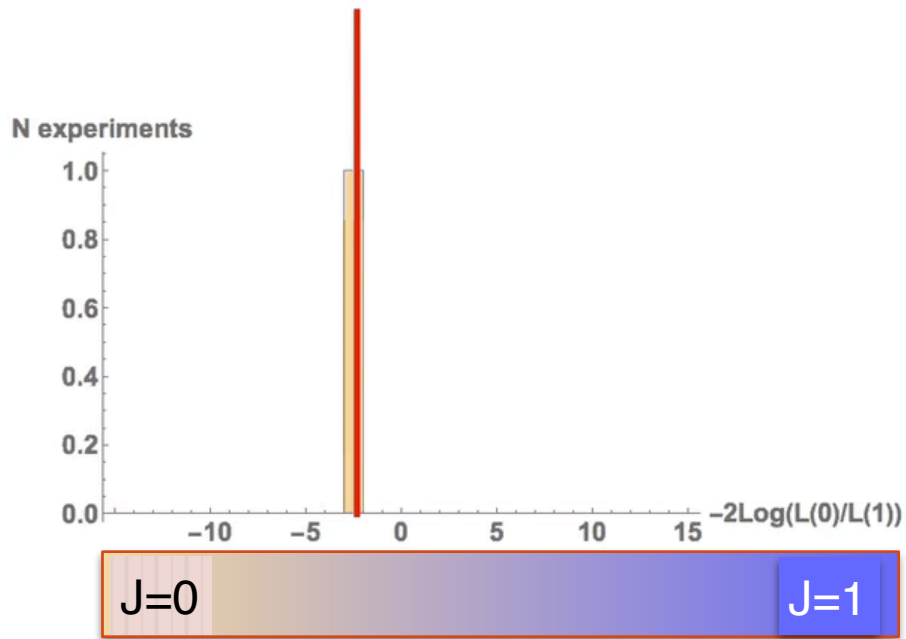
$$\alpha = \text{prob}(\lambda \leq \eta)$$



Building PDF

Build the pdf of the test statistic

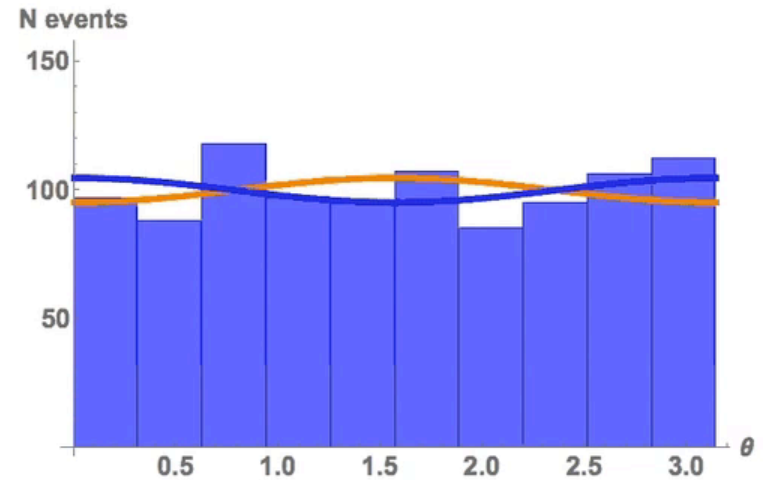
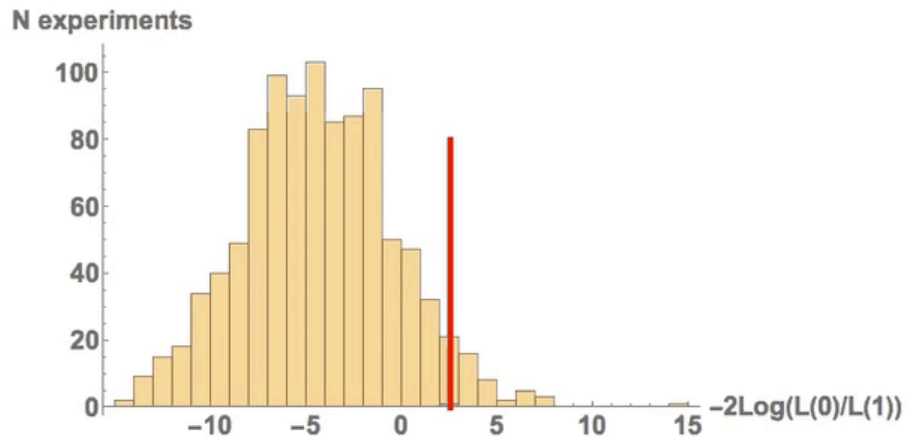
$$q_{NP} = q_{NP}(x) = -2 \ln \frac{L(H_0 | x)}{L(H_1 | x)}$$



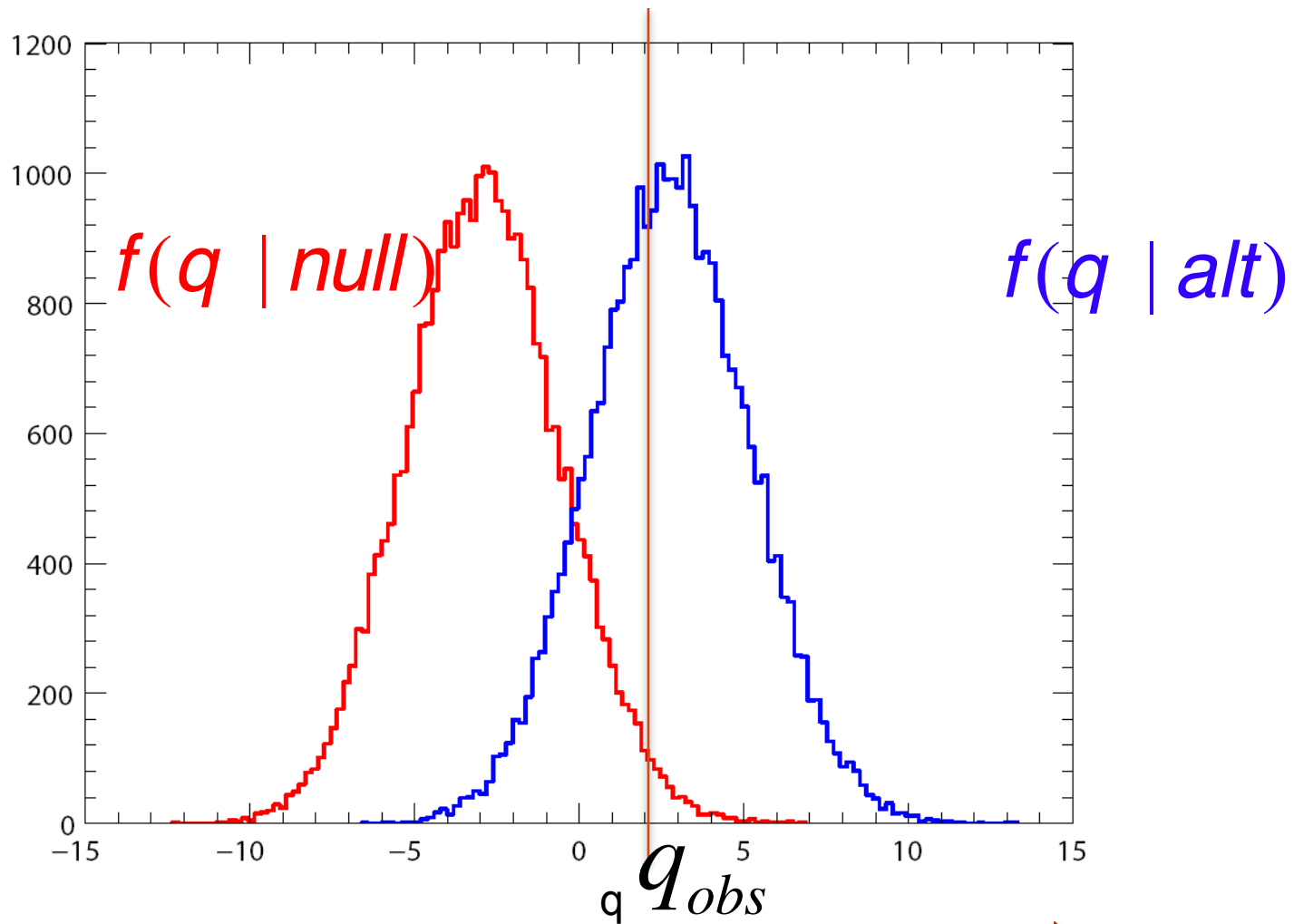
Building PDF

Build the pdf of the test statistic

$$q_{NP} = q_{NP}(x) = -2 \ln \frac{L(H_0 | x)}{L(H_1 | x)}$$



PDF of a test statistic



Null like

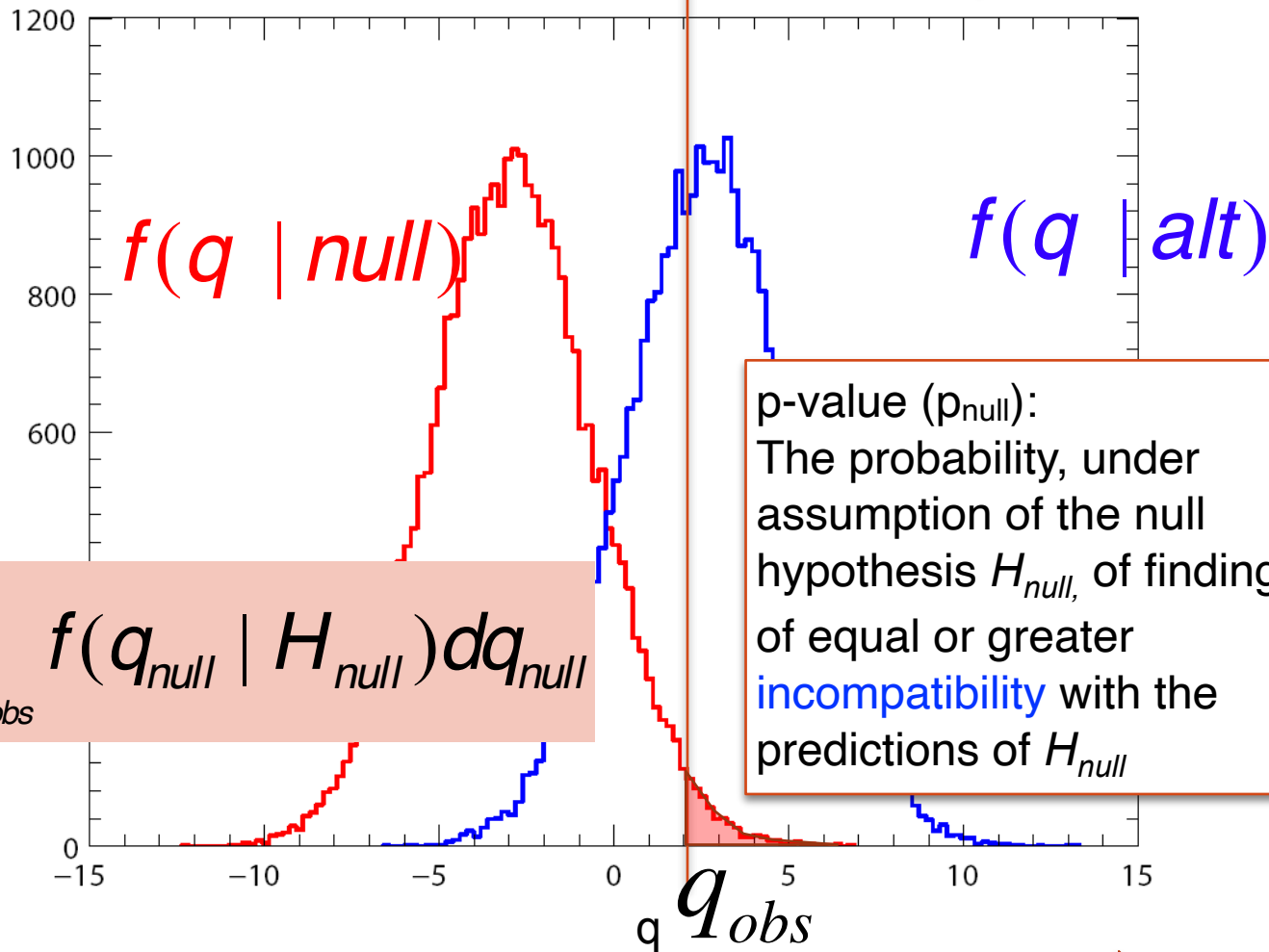


alt like



PDF of a test statistic

If $p \leq \alpha$ reject null



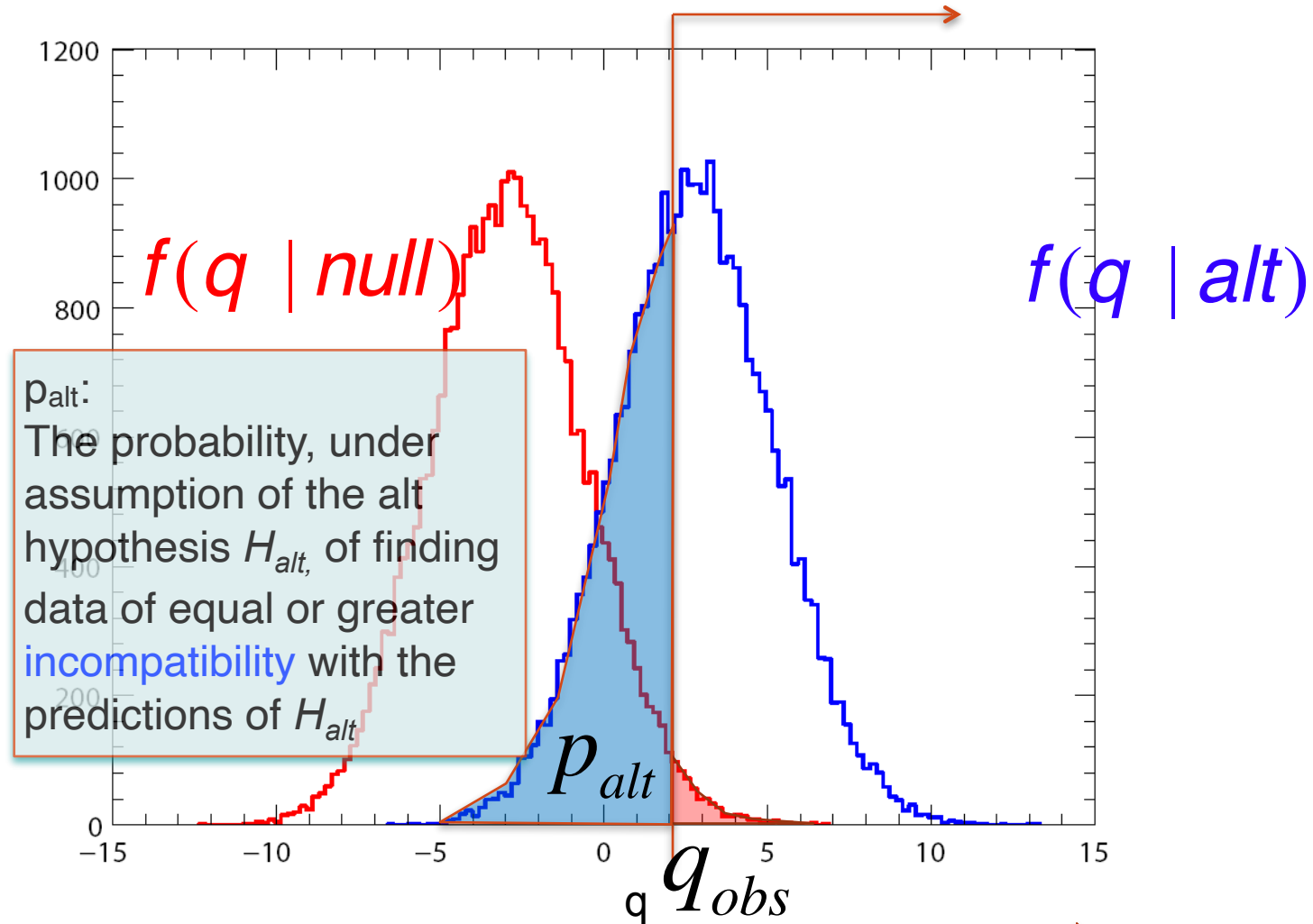
$$p = \int_{q_{obs}}^{\infty} f(q_{null} \mid H_{null}) dq_{null}$$

p-value (p_{null}):
The probability, under assumption of the null hypothesis H_{null} , of finding data of equal or greater incompatibility with the predictions of H_{null}



PDF of a test statistic

If $p \leq \alpha$ reject null



Null like

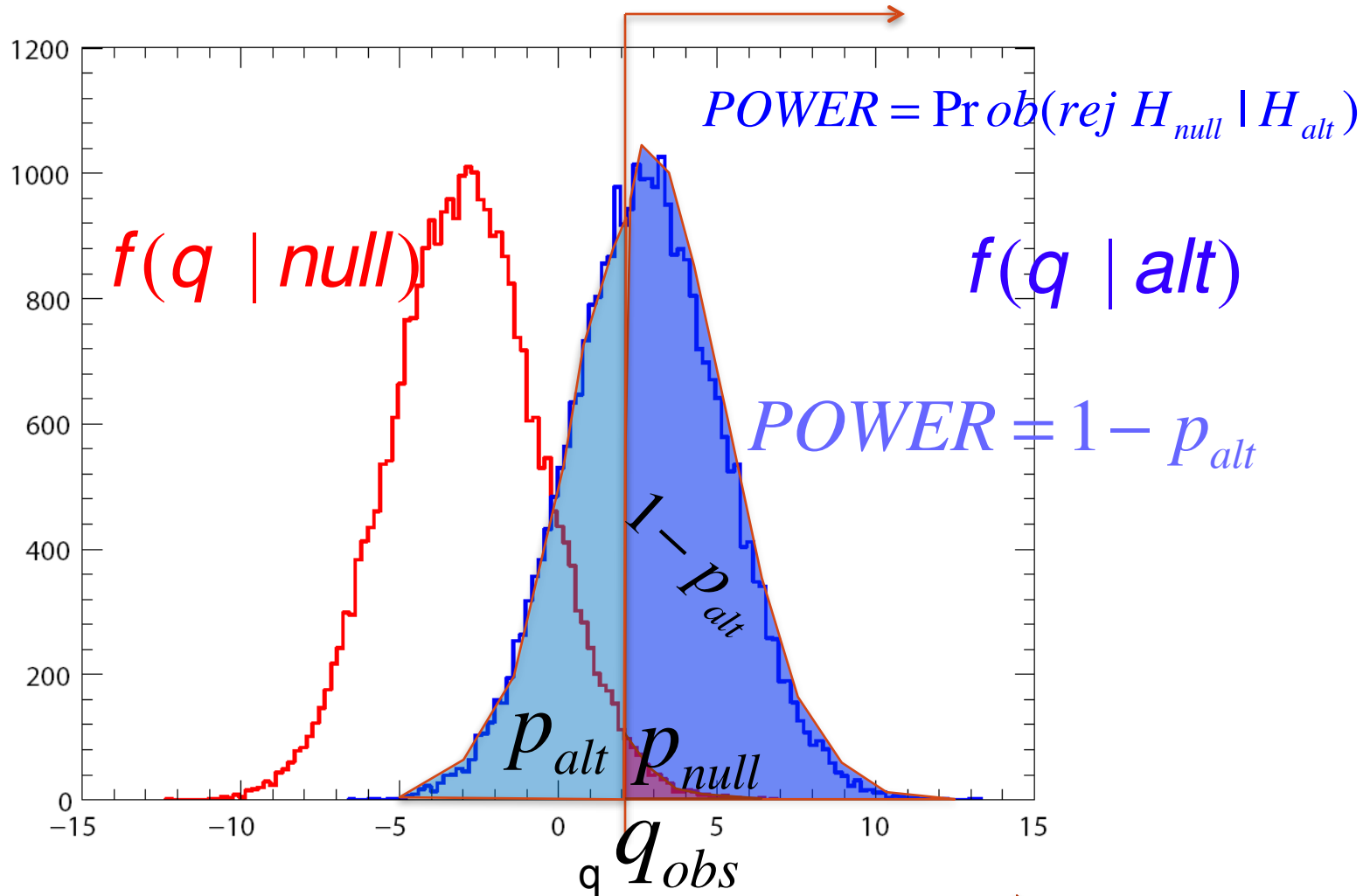


alt like



PDF of a test statistic

If $p \leq \alpha$ reject null



Null like



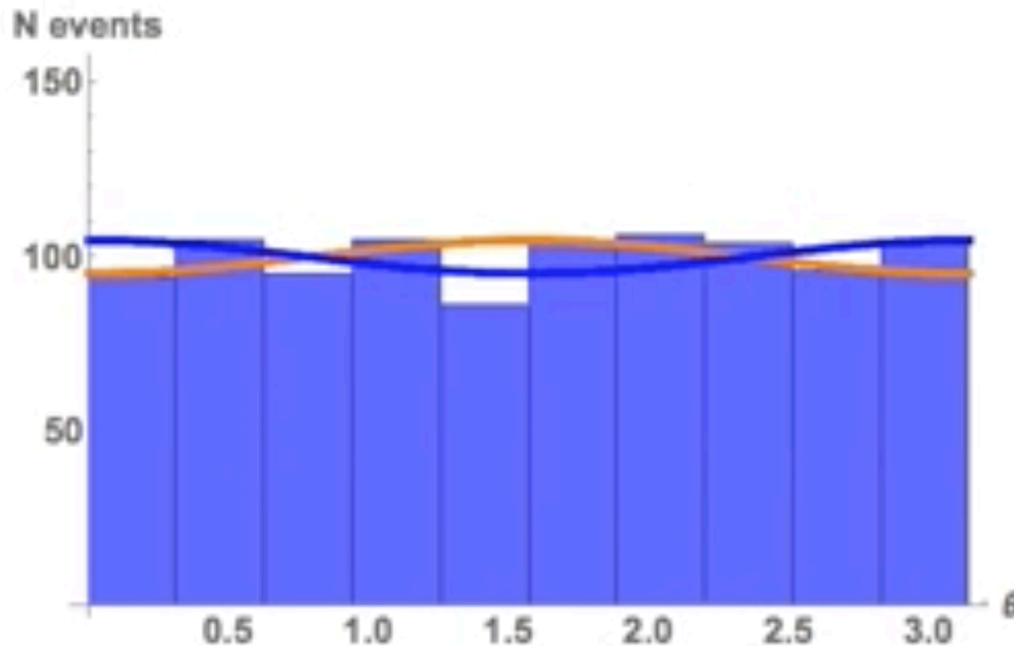
alt like



Power and Luminosity

For a given significance the power increases with increased luminosity

Luminosity \sim Total number of events in an experiment

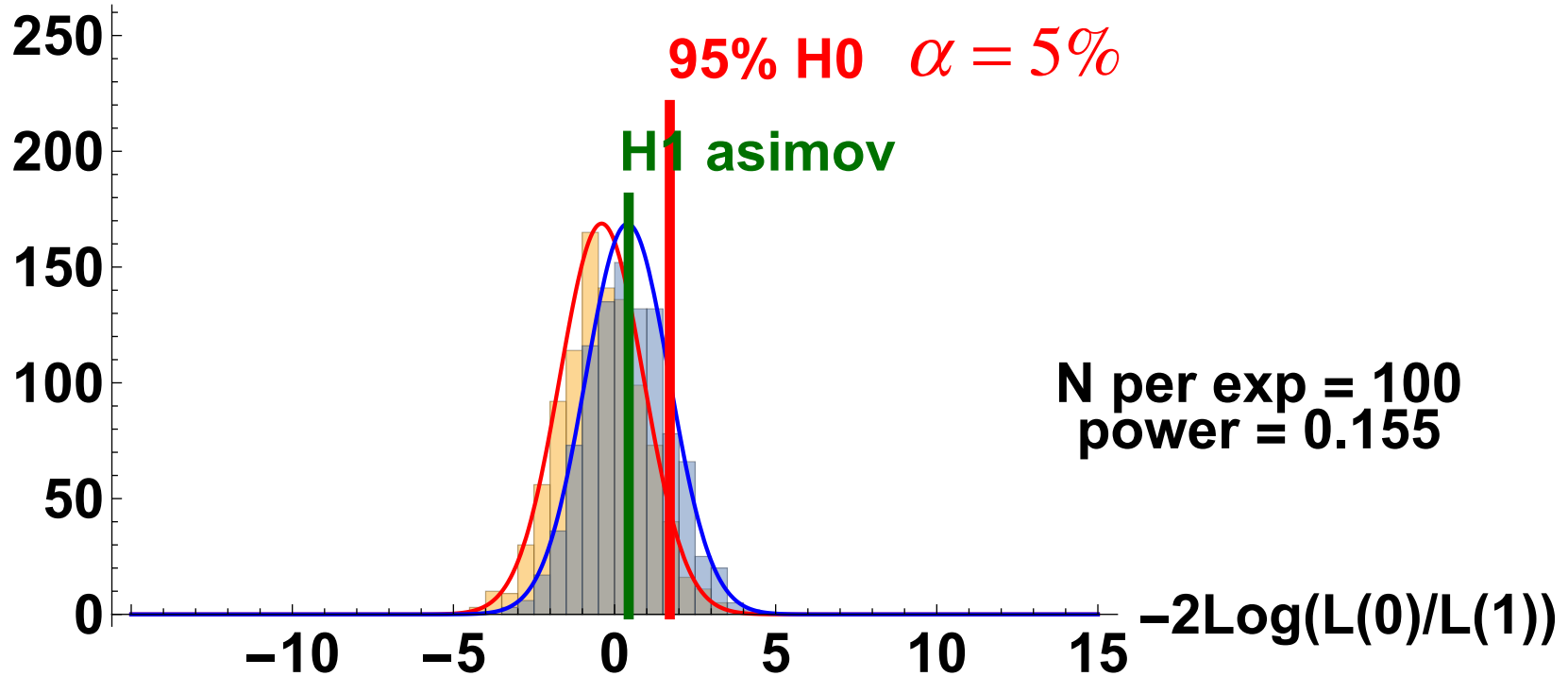


Hard to tell $f(q|J=0)$ from $f(q|J=1)$

→ CLs

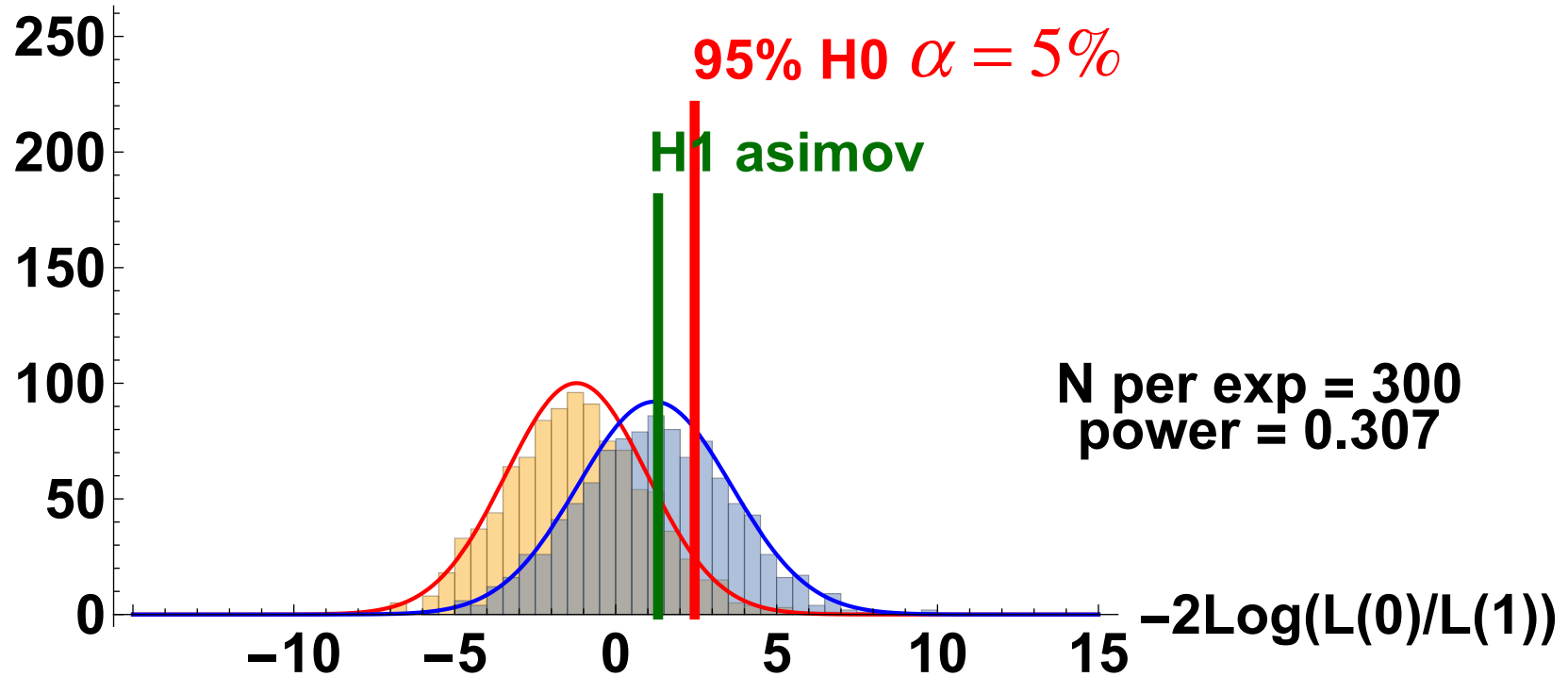
$$POWER = \text{Prob}(\text{rej } H_{null} | H_{alt})$$

N experiments



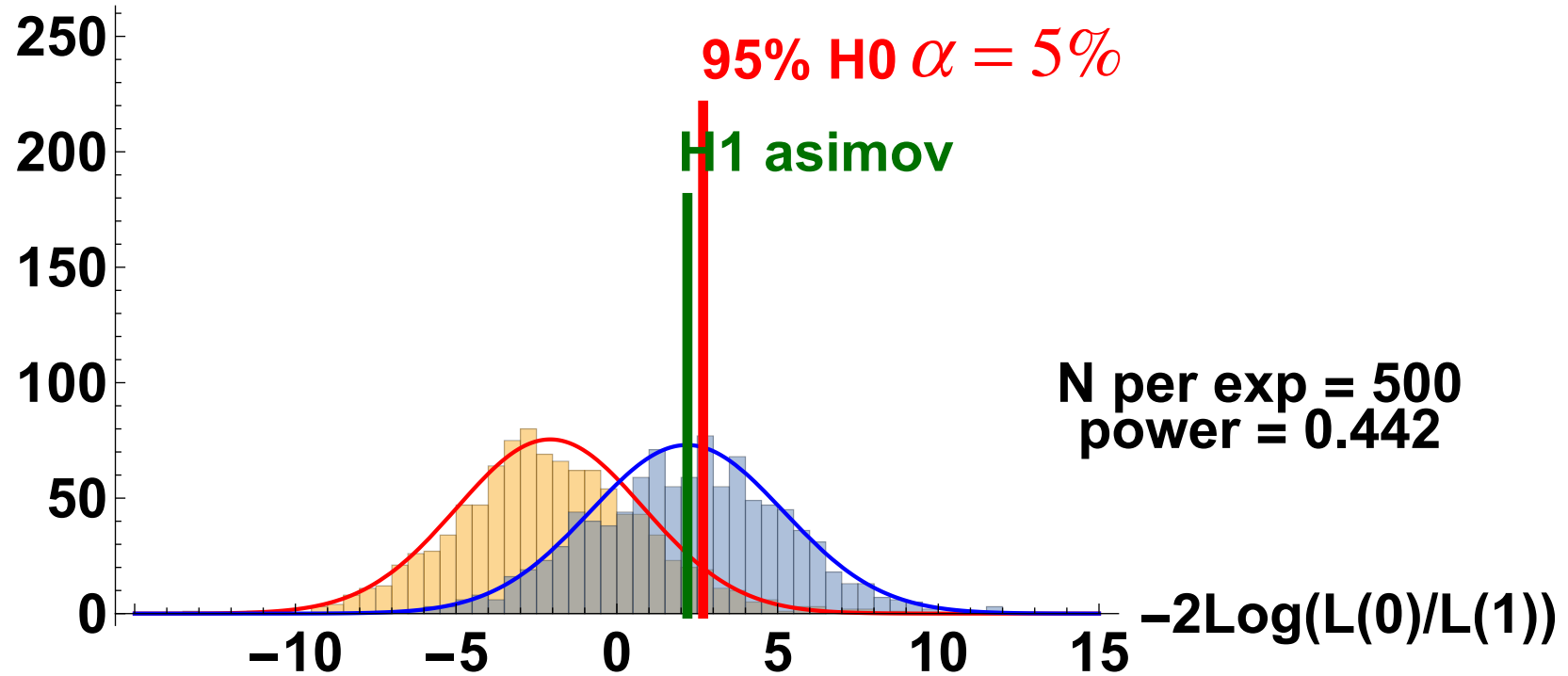
N experiments

$$POWER = \text{Prob}(\text{rej } H_{null} | H_{alt})$$



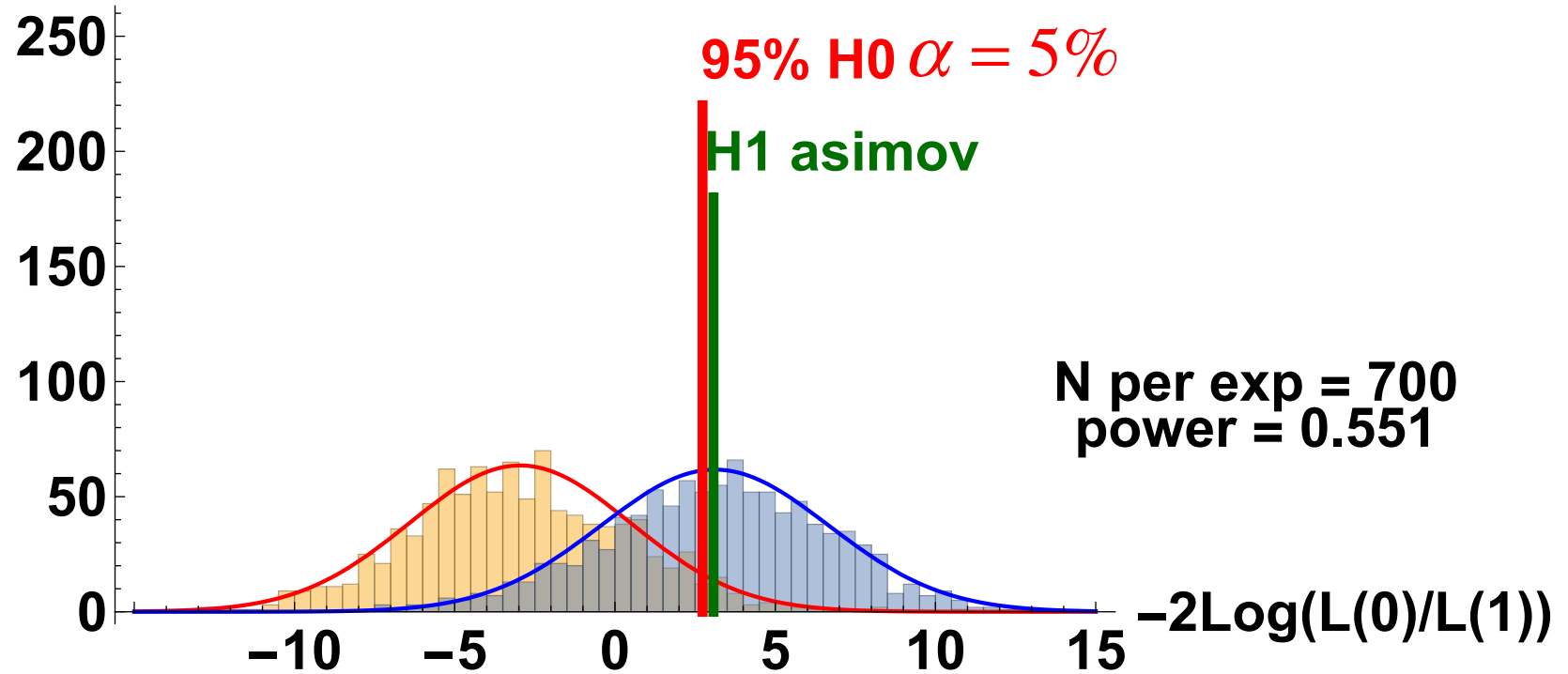
N experiments

$$POWER = \text{Prob}(\text{rej } H_{null} | H_{alt})$$



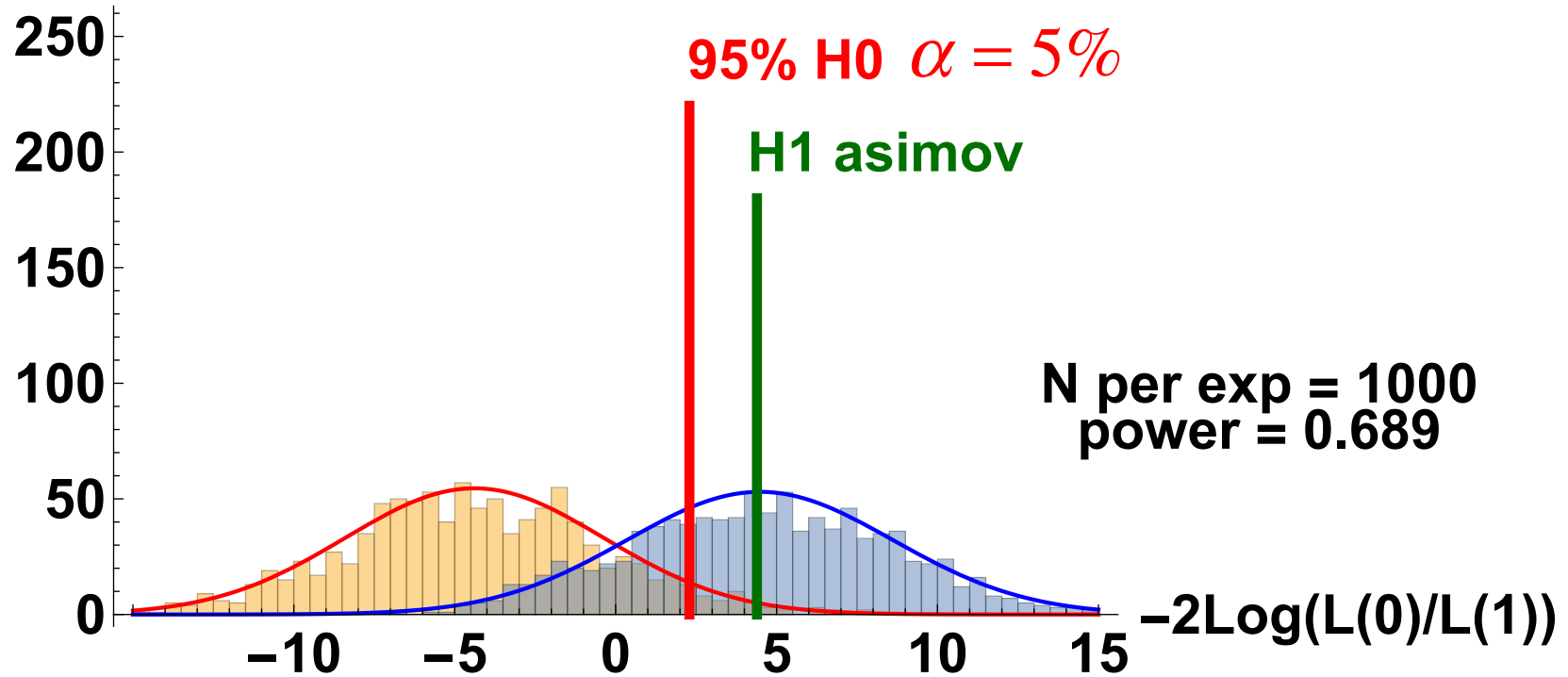
N experiments

$$POWER = \text{Pr ob}(rej H_{null} | H_{alt})$$



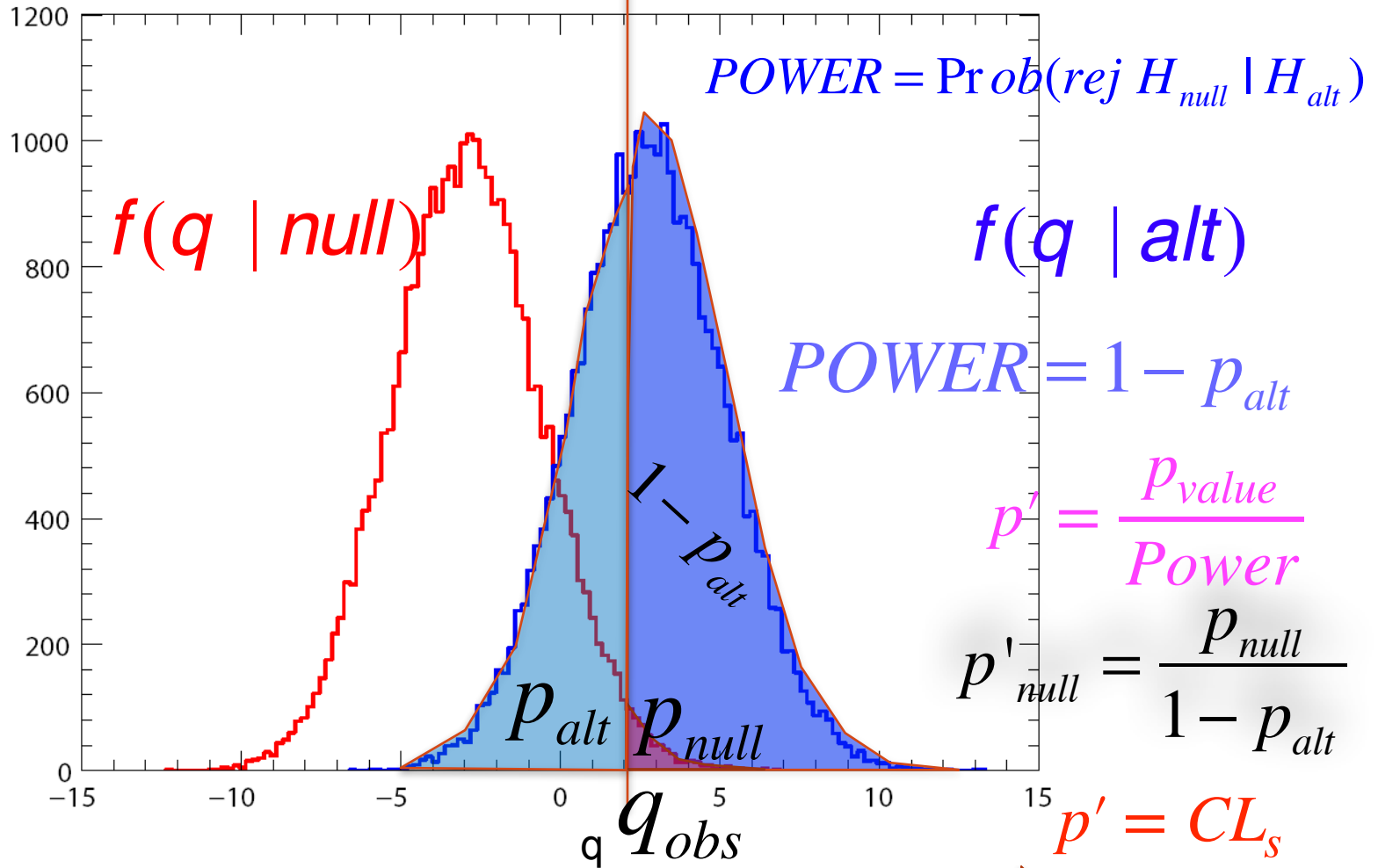
N experiments

$$POWER = \text{Prob}(\text{rej } H_{null} | H_{alt})$$



CLs

If $p \leq \alpha$ reject null



Null like



alt like

$p' = CL_s$



Test Spin 0 parity

$$H_0 = 0^+$$

$$H_1 = 0^-$$

$$q^{NP} = -2 \ln \frac{L(H_0)}{L(H_1)}$$

$$p_{H_1}(\text{exp} | H_0) = 0.37\%$$

$$p_{H_1}(\text{obs}) = 1.5\%$$

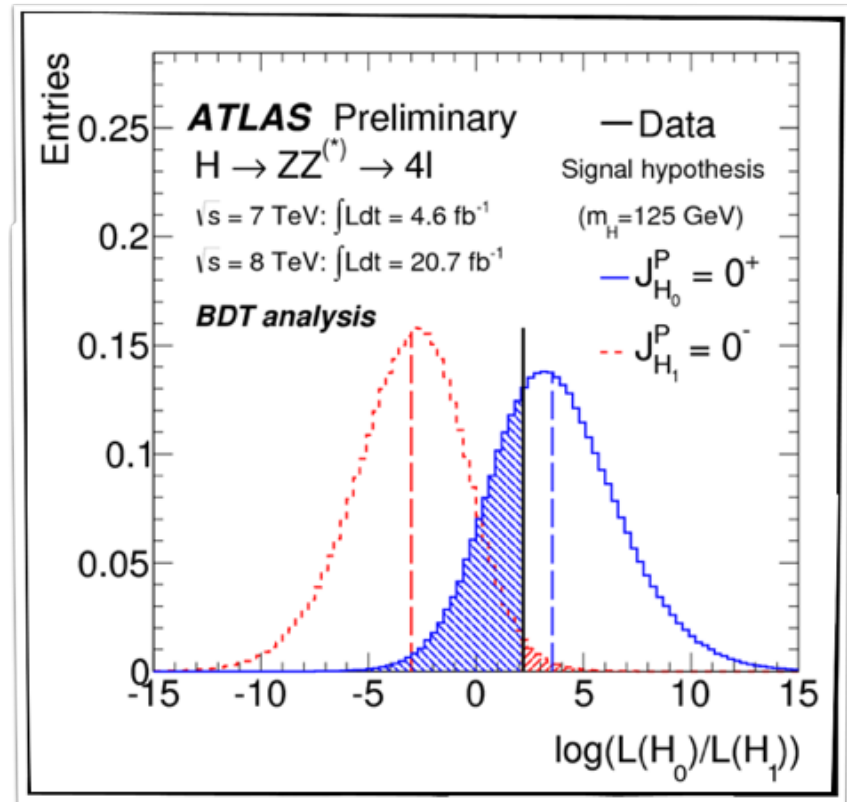
$$p_{H_0}(\text{obs}) = 31\%$$

$$p_{H_1}^{CL_s}(\text{obs}) = 2.2\%$$

$$p_{H_1}^{CL_s} = \frac{p_{H_1}}{1 - p_{H_0}} = \frac{1.5\%}{1 - 0.31} = 2.2\%$$

Which means

$J^P=0^-$ is excluded at the 97.8% CL in favour of $J^P=0^+$



H_1 like

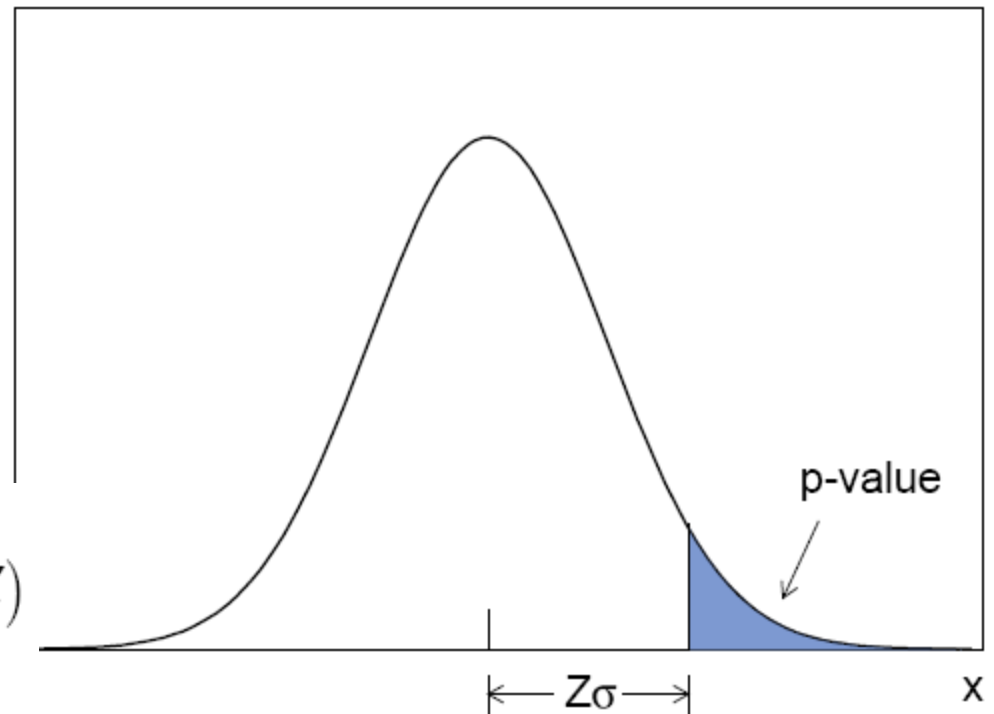
H_0 like

From p-values to Gaussian significance

It is a custom to express the p-value as the significance associated to it, had the pdf were Gaussians

$$p = \int_Z^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1 - \Phi(Z)$$

$$Z = \Phi^{-1}(1 - p)$$



A significance of $Z = 5$ corresponds to $p = 2.87 \times 10^{-7}$

Beware of 1 vs 2-sided definitions!

p-value – testing the null hypothesis

When testing the b hypothesis (null= b), it is custom to set $\alpha = 2.9 \cdot 10^{-7}$

→ if $p_b < 2.9 \cdot 10^{-7}$ the b hypothesis is rejected

→ Discovery

When testing the $s+b$ hypothesis (null= $s+b$), set $\alpha = 5\%$
if $p_{s+b} < 5\%$ the signal hypothesis is rejected at the 95%
Confidence Level (CL)

→ Exclusion

Profile Likelihood with Nuisance Parameters

$$q_{\mu} = -2 \ln \frac{L(\mu s + \hat{b}_{\mu})}{L(\hat{\mu} s + \hat{b})}$$

$$q_{\mu} = -2 \ln \frac{\max_b L(\mu s + b)}{\max_{\mu, b} L(\mu s + b)}$$

$$q_{\mu} = q_{\mu}(\hat{\mu}) = -2 \ln \frac{L(\mu s + \hat{b}_{\mu})}{L(\hat{\mu} s + \hat{b})}$$

$\hat{\mu}$ MLE of μ

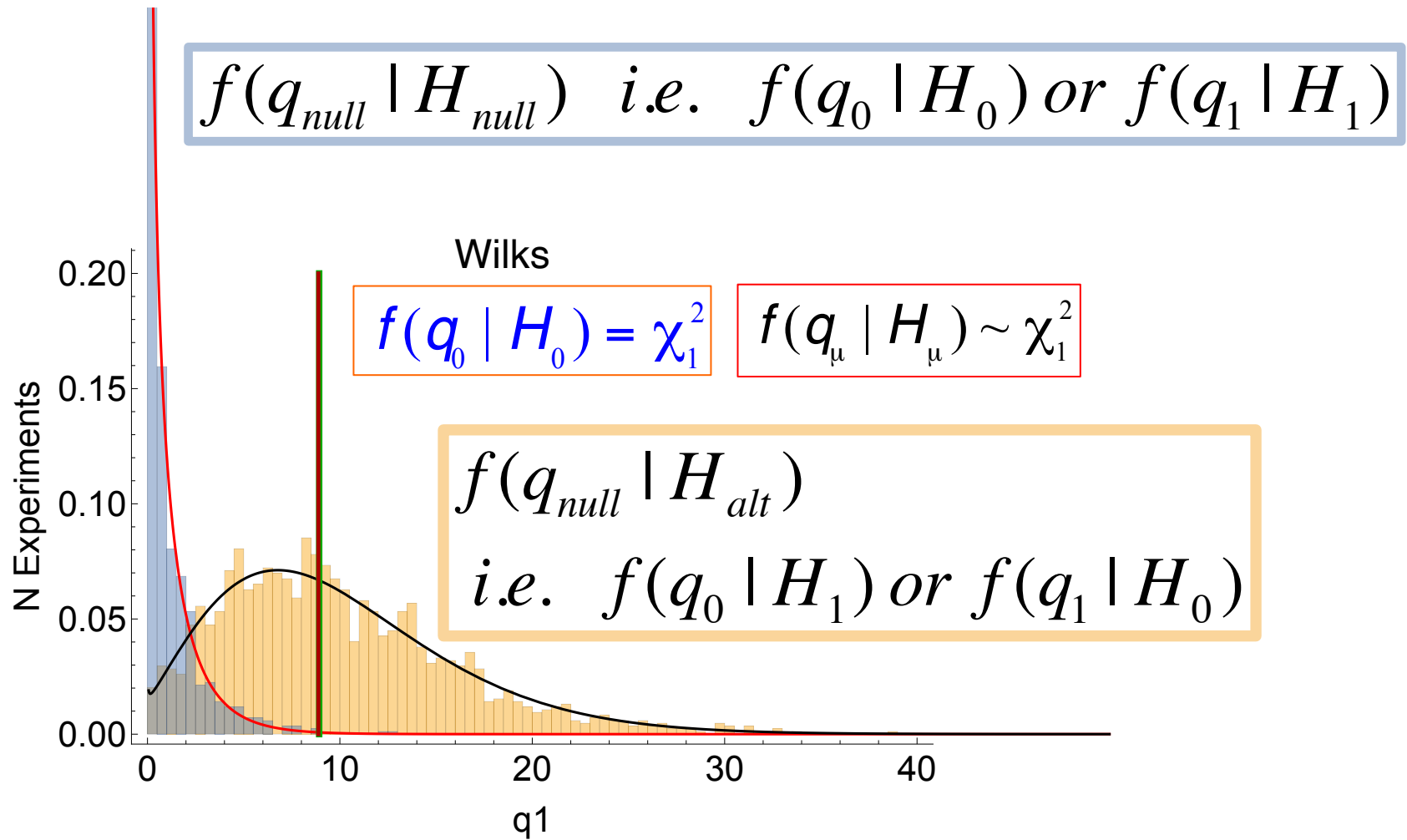
\hat{b} MLE of b

\hat{b}_{μ} MLE of b fixing μ

$\hat{\theta}_{\mu}$ MLE of θ fixing μ



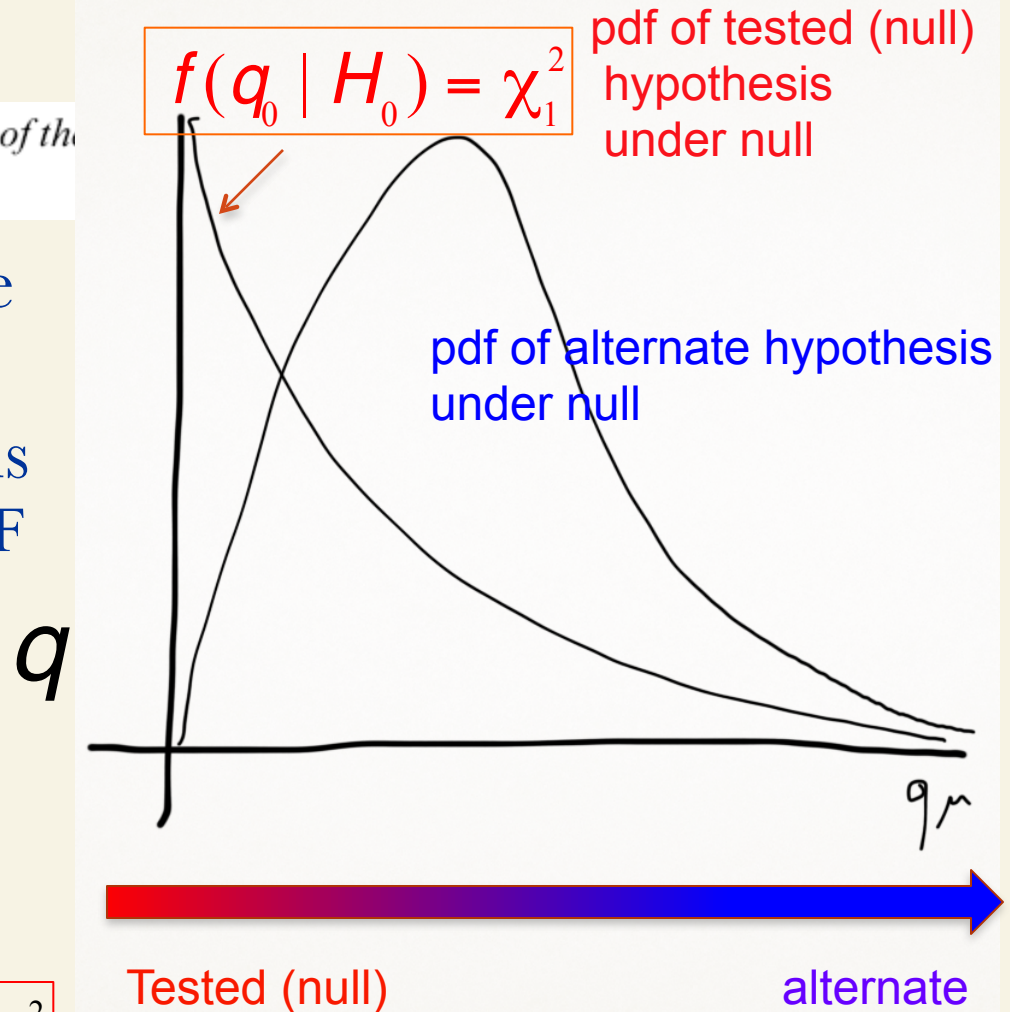
This Lecture's Questions



Wilks Theorem

S.S. Wilks, *The large-sample distribution of the*
Ann. Math. Statist. **9** (1938) 60-2.

- *Wilks' theorem says that the pdf of the statistic under the null hypothesis approaches a chi-square PDF for one degree of freedom*



$$f(q_0 | H_0) = \chi_1^2$$

$$f(q_\mu | H_\mu) \sim \chi_1^2$$



Classification of Test Statistics

Test Stat.	Purpose	Expression	LR
q_0	discovery of positive signal	$q_0 = \begin{cases} -2 \ln \lambda(0) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases}$	$\lambda(0) = \frac{L(0, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\hat{\theta}})}$
t_μ	2-sided measurement	$t_\mu = -2 \ln \lambda(\mu)$	$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\hat{\theta}})}$
\tilde{t}_μ	avoid negative signal (Feldman-Cousins)	$\tilde{t}_\mu = -2 \ln \tilde{\lambda}(\mu)$	$\tilde{\lambda}(\mu) = \begin{cases} \frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(\hat{\mu}, \hat{\hat{\theta}})} & \hat{\mu} \geq 0 \\ \frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(0, \hat{\hat{\theta}}(0))} & \hat{\mu} < 0 \end{cases}$
q_μ	exclusion	$q_\mu = \begin{cases} -2 \ln \lambda(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases}$	
\tilde{q}_μ	exclusion of positive signal	$\tilde{q}_\mu = \begin{cases} -2 \ln \frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(0, \hat{\hat{\theta}}(0))} & \hat{\mu} < 0, \\ -2 \ln \frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(\hat{\mu}, \hat{\hat{\theta}})} & 0 \leq \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases}$	



Study Case 2: Bump Hunt



Bump Hunt

Test H_0 with q_0 , Reject $H_0 \Rightarrow Discovery$

$$q_0 = \begin{cases} -2 \ln \lambda(0) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases} \quad \lambda(0) = \frac{L(0, \hat{\theta}_0)}{L(\hat{\mu}, \hat{\theta})}$$

Test $H_\mu(m_H)$ with q_μ Reject $H_\mu(m_H) \Rightarrow$

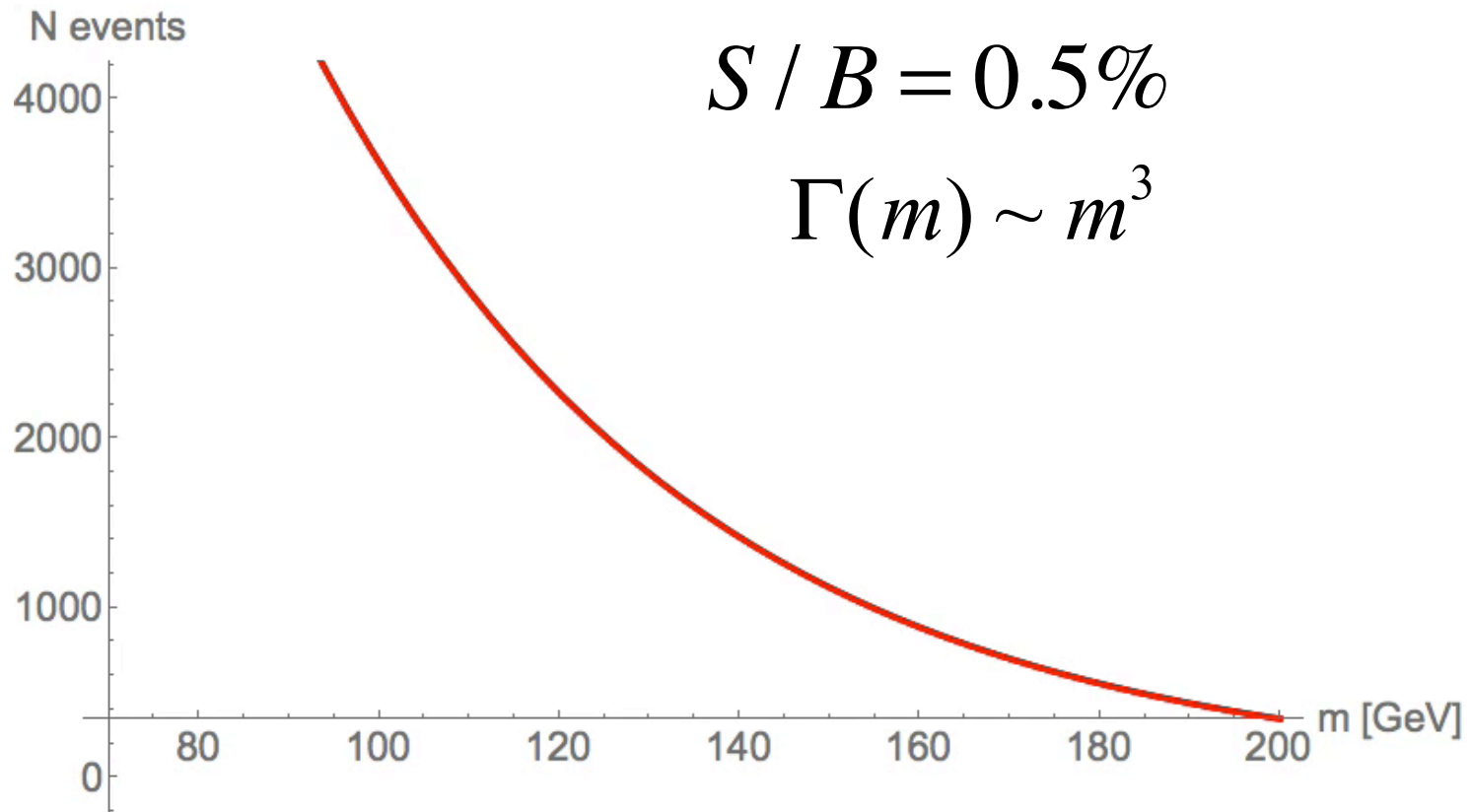
Exclusion of a Higgs with $m_H \Rightarrow \mu_{up}(m_H)$

$$q_\mu = \begin{cases} -2 \ln \lambda(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} \quad \lambda(\mu) = \frac{L(\mu, \hat{\theta}_\mu)}{L(\hat{\mu}, \hat{\theta})}$$

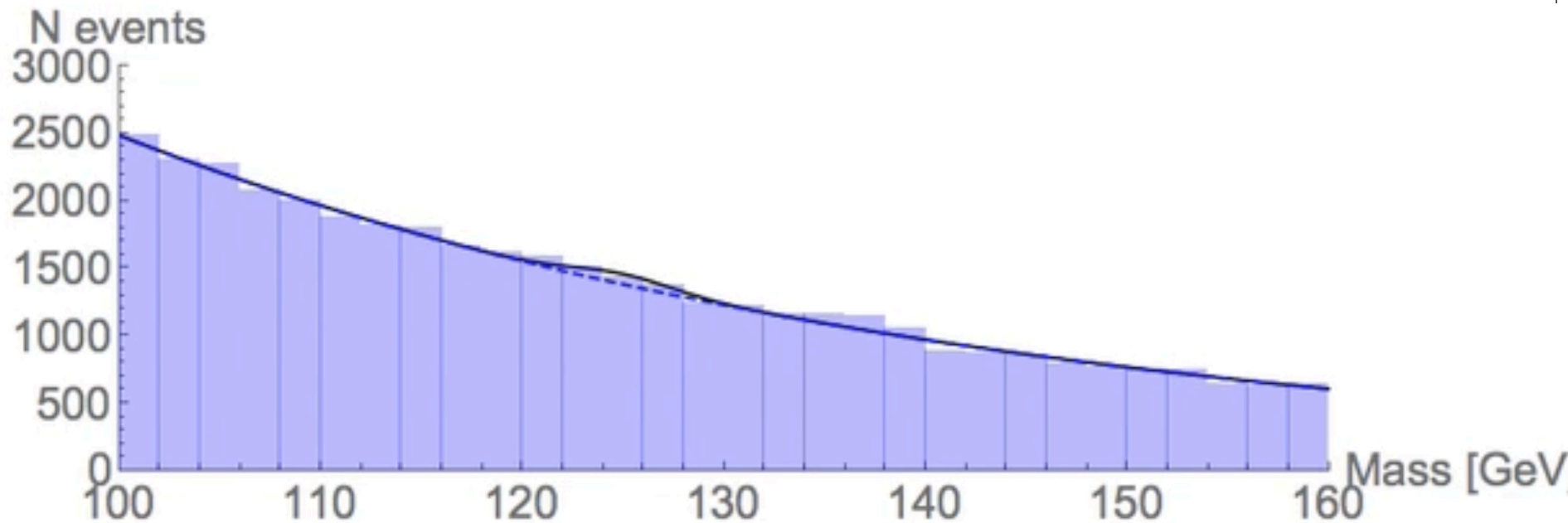


Bump Hunt

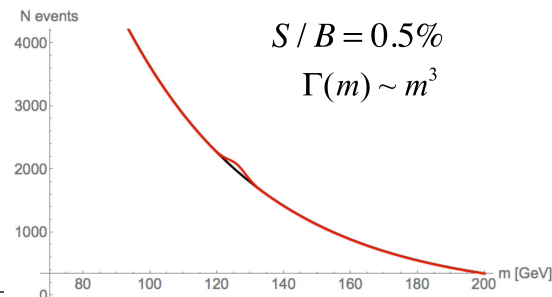
Gamma Gamma like BG and a Gaussian signal on top of it



A GammaGammaLike Signal



Luminosity is the number of events in the histogram



Asymptotic Approximation

CCGV

Asymptotic formulae for likelihood-based tests of new physics

Glen Cowan (Royal Holloway, U. of London), Kyle Cranmer (New York U.), Eilam Gross, Ofer Vitells (Weizmann Inst.). Jul 10, 2010. 25 pp.

Published in *Eur.Phys.J. C71 (2011) 1554*, Erratum: *Eur.Phys.J. C73 (2013) 2501*



Kyle
Cranmer

Glen
Cowan

Ofer
Vitells

E.G.



Test Statistic

$$t_{\mu} = -2\ln\lambda(\mu)$$

$$t_{\mu} = -2\ln\lambda(\mu) \quad \lambda(\mu) = \frac{L(\mu, \hat{\theta}_{\mu})}{L(\hat{\mu}, \hat{\theta})}$$

Higher values of t_{μ} correspond to increasing incompatibility between the data and μ



Wald Theorem

$$\lambda(\mu) = \frac{L(\mu, \hat{\theta}_\mu)}{L(\hat{\mu}, \hat{\theta})} \quad t_\mu = -2 \ln \lambda(\mu) \quad \text{Wilks} \Rightarrow f(t_\mu | \mu) \sim \chi_1^2$$

How does t_μ distribute under $H_{\mu'}$ ($\mu' \neq \mu$)

A. Wald, *Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large*, Transactions of the American Mathematical Society, Vol. 54, No. 3 (Nov., 1943), pp. 426-482.

$$t_\mu = -2 \ln \lambda(\mu) = \frac{(\mu - \hat{\mu})^2}{\sigma_{\hat{\mu}}^2} + O(1/\sqrt{N})$$

(Use the Asimov Dataset to estimate σ)

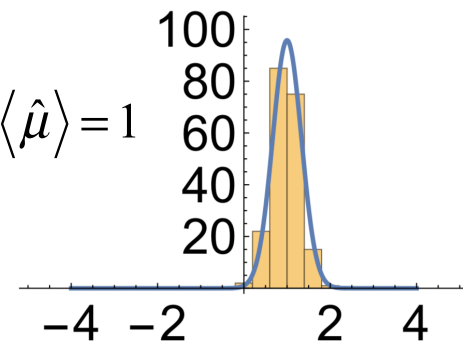
$f(t_\mu | \mu')$ follows a noncentral Chi squared distribution

with non-centrality parameter $\Lambda = \frac{(\mu - \mu')^2}{\sigma^2}$ with 1 d.o.f

where $\hat{\mu} \sim G(\mu', \sigma)$

N is the sample size

$$\mu' = 1 \Rightarrow \langle \hat{\mu} \rangle = 1$$



Wald Theorem

$$t_{\mu} = -2 \ln \lambda(\mu) = \frac{(\mu - \hat{\mu})^2}{\sigma_{\hat{\mu}}^2} + O(1/\sqrt{N})$$

$$\hat{\mu} \sim G(\mu', \sigma)$$

N is the sample size

$f(t_{\mu} | \mu')$ follows a noncentral Chi squared distribution

with non-centrality parameter $\Lambda = \frac{(\mu - \mu')^2}{\sigma^2}$ with 1 d.o.f

$$f(t_{\mu}; \Lambda) = \frac{1}{2\sqrt{t_{\mu}}} \frac{1}{\sqrt{2\pi}} \left[\exp\left(-\frac{1}{2}(\sqrt{t_{\mu}} + \sqrt{\Lambda})^2\right) + \exp\left(-\frac{1}{2}(\sqrt{t_{\mu}} - \sqrt{\Lambda})^2\right) \right]$$

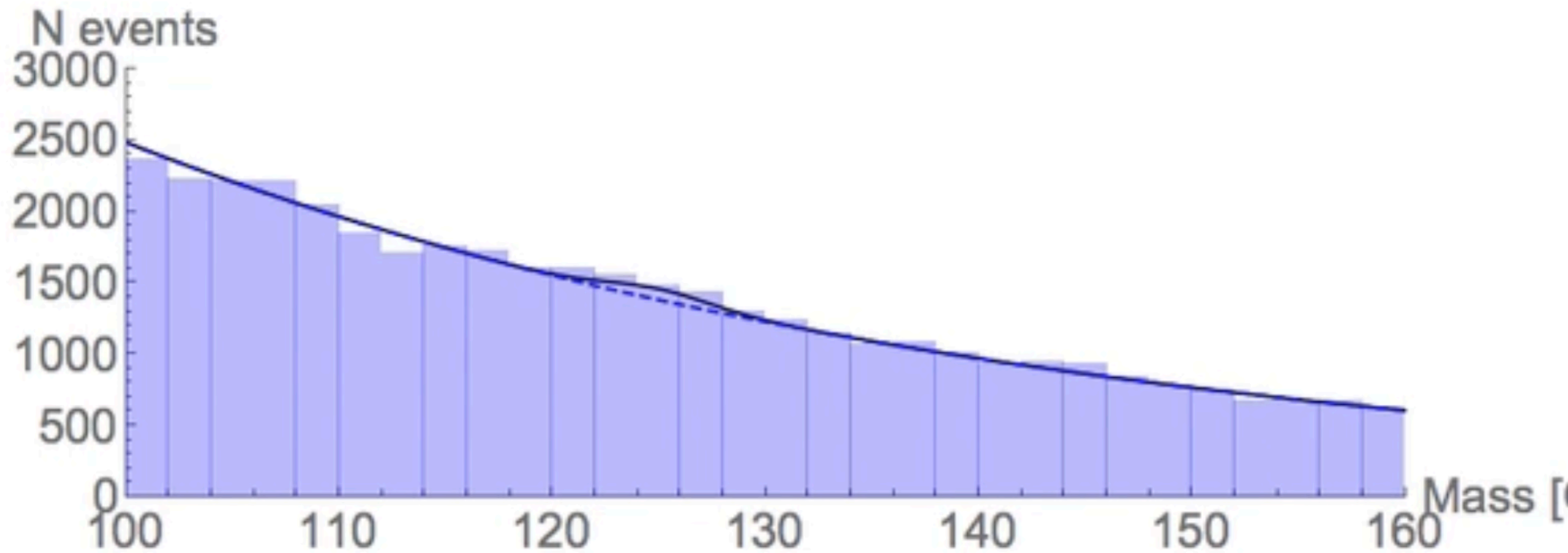
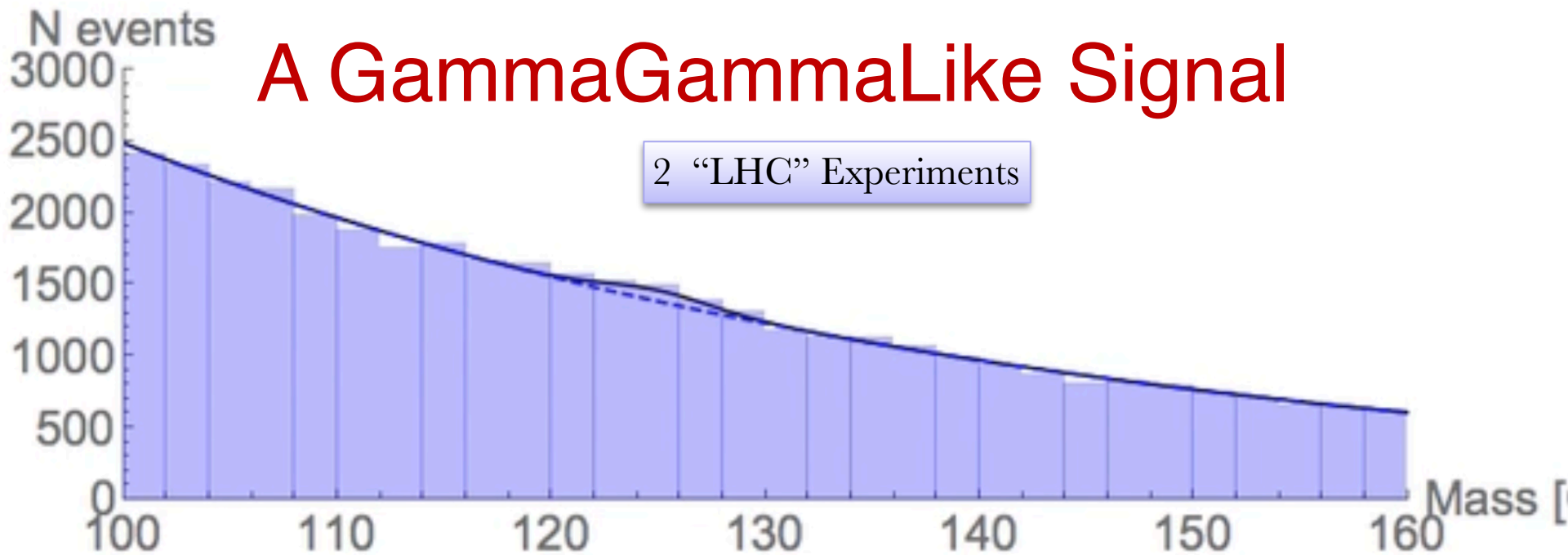
for $\mu' = \mu$ we retrieve Wilks' theorem

$$f(t_{\mu}) = \frac{1}{\sqrt{2\pi t_{\mu}}} e^{-\frac{1}{2}t_{\mu}} = \chi^2$$

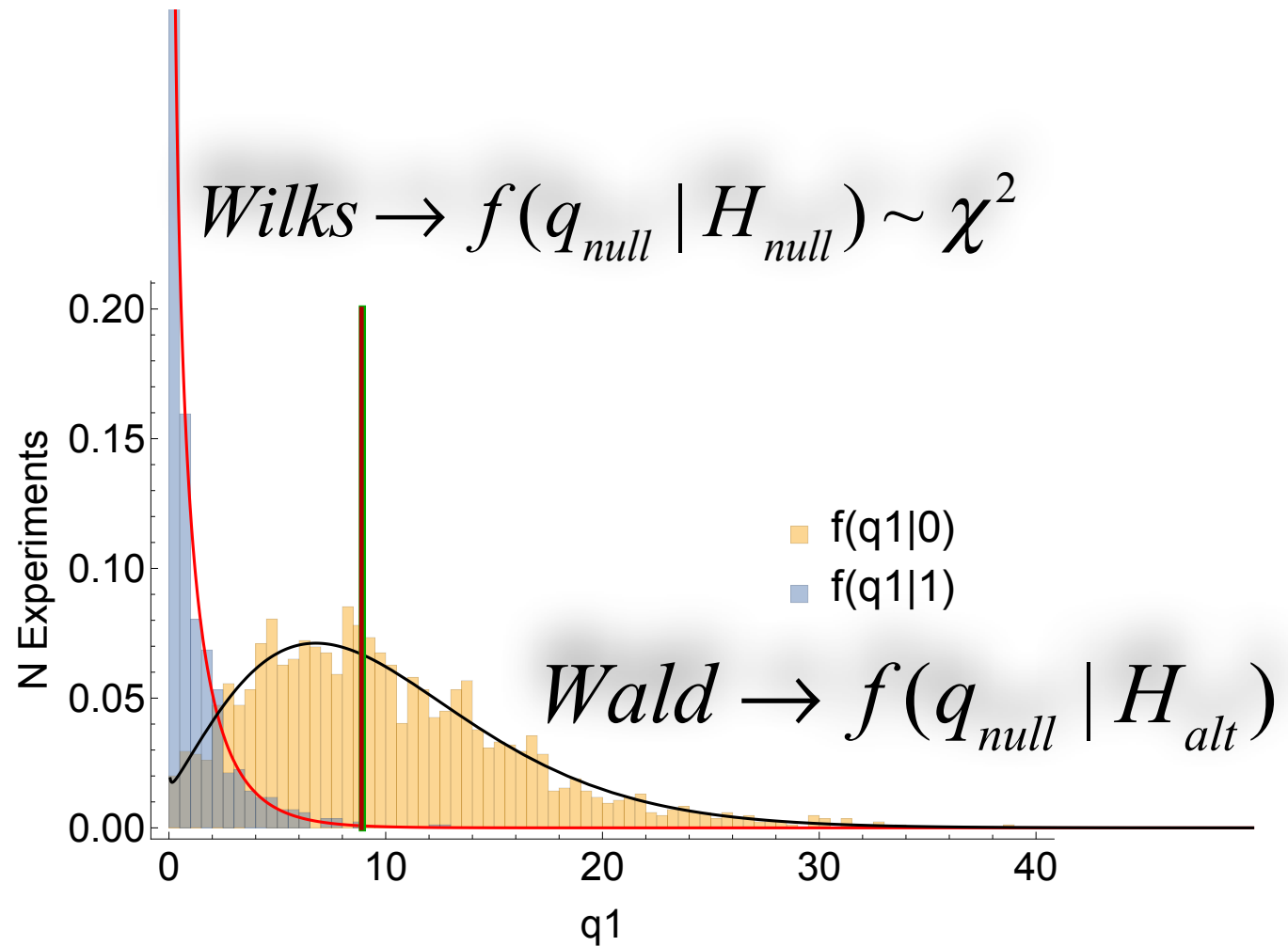


A GammaGammaLike Signal

2 "LHC" Experiments



Asymptotics

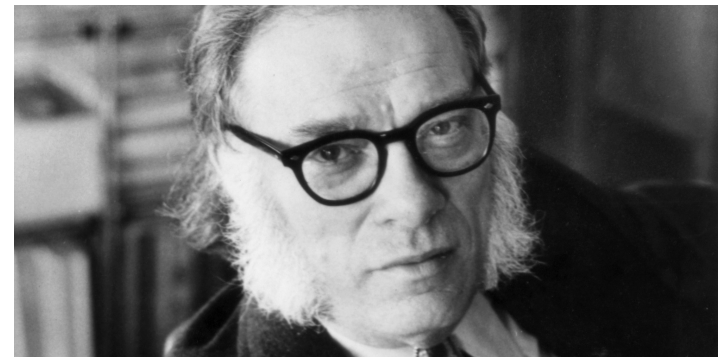
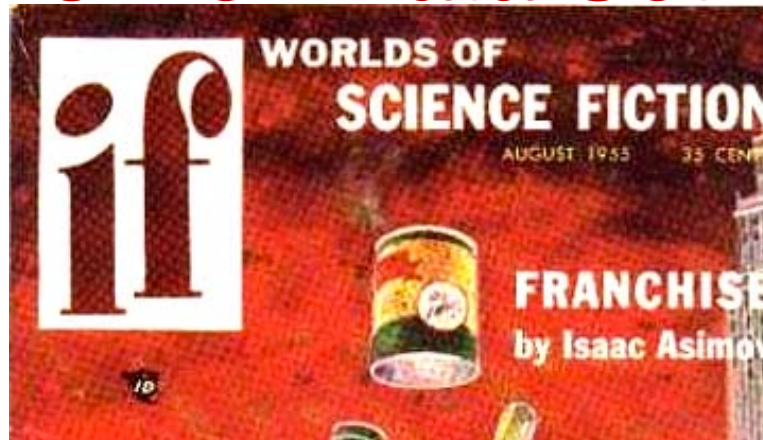


Estimating the Sensitivity of an Experiment

- Estimate the expected significance one could achieve (for discovering the Higgs Boson) with a given analysis, a given Luminosity and CM energy..
- **Option 1:**
 - Toss, say, 1,000,000 BG only events (null) and derive the BG-only pdf of q , $f(q_{\text{null}}|BG)$.
Toss another 1,000,000 S+BG (alt) events and find the significance for each one of them
then, find the median significance....
 - This may take ages..., is there a shortcut?
- **Option 2:**
 - Asymptotics+Asimov Data Set

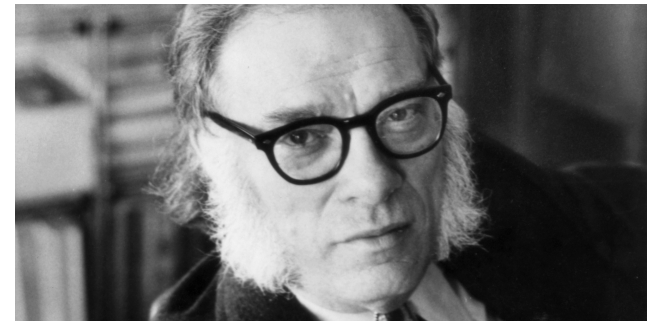


The Asimov Data Set



In the future, the United States has converted to an "electronic democracy" where the computer Multivac selects a single person to answer a number of questions. Multivac will then use the answers and other data to determine what the results of an election would be, avoiding the need for an actual election to be held.

The Asimov Data Set



- *The use of a single representative individual to stand in for the entire population can help in evaluating the sensitivity of a statistical method.*
- *The "Asimov data set": an ensemble of simulated experiments can be replaced by a single representative one.*

Estimating the Sensitivity of an Experiment

- one can replace each ensemble of the alternate-hypothesis experiments with one data set that represents the typical experiment.

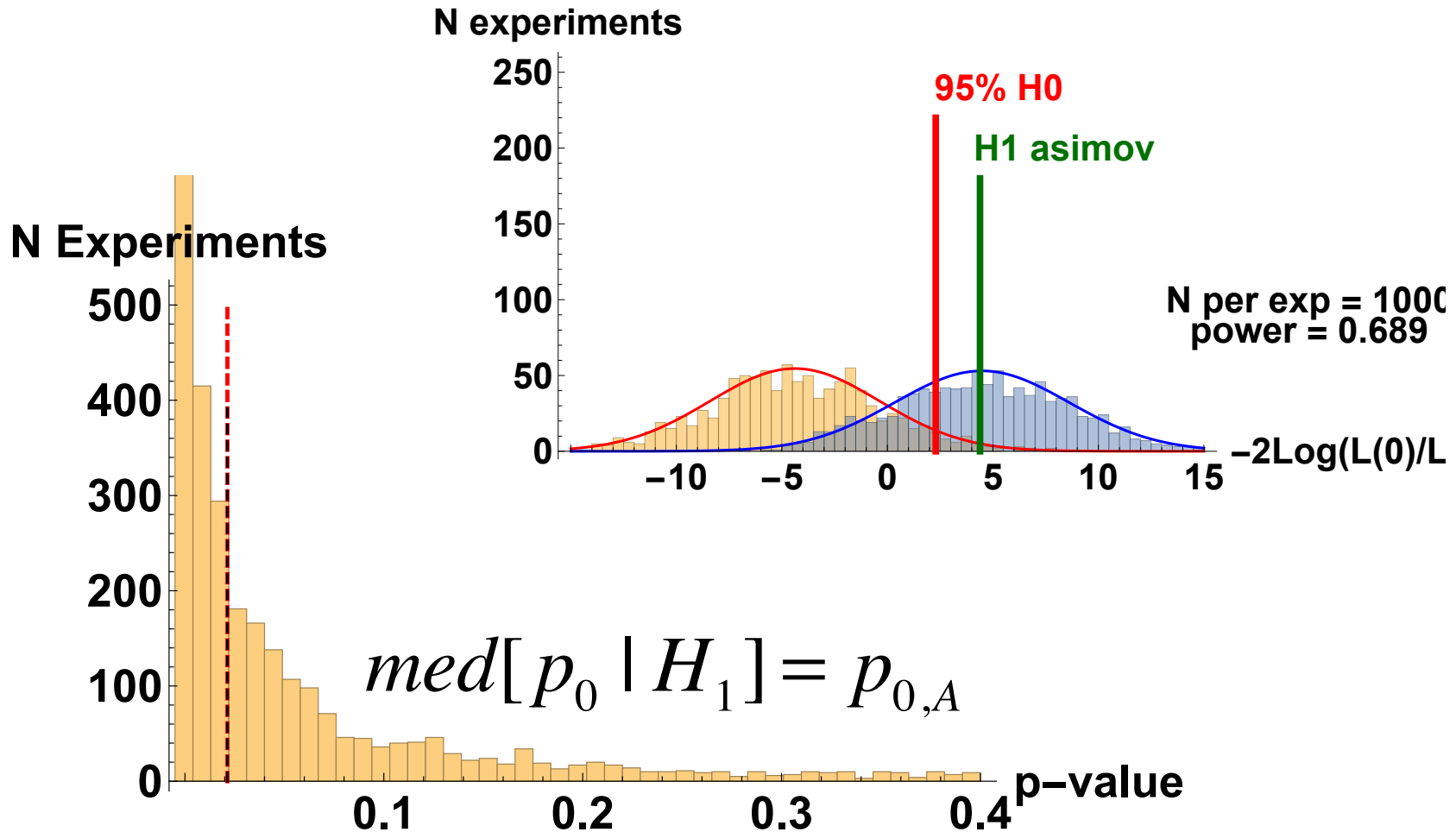
This “Asimov” data set delivers the desired median sensitivity. Hence, one is exempted from the need to perform an ensemble of experiments for each set of parameters.

- The Asimov data set is constructed such that when one uses it to evaluate the estimators for all parameters, one obtains the true parameter values.
- the Asimov data set can trivially be constructed from the true parameters values. For example, a set corresponding to the H_1 hypothesis is $n_A = s + b$. and the one correspond to the H_0 hypothesis is $n_A = b$.
- As strange as it reads, the Asimov data set is not necessarily an integer.



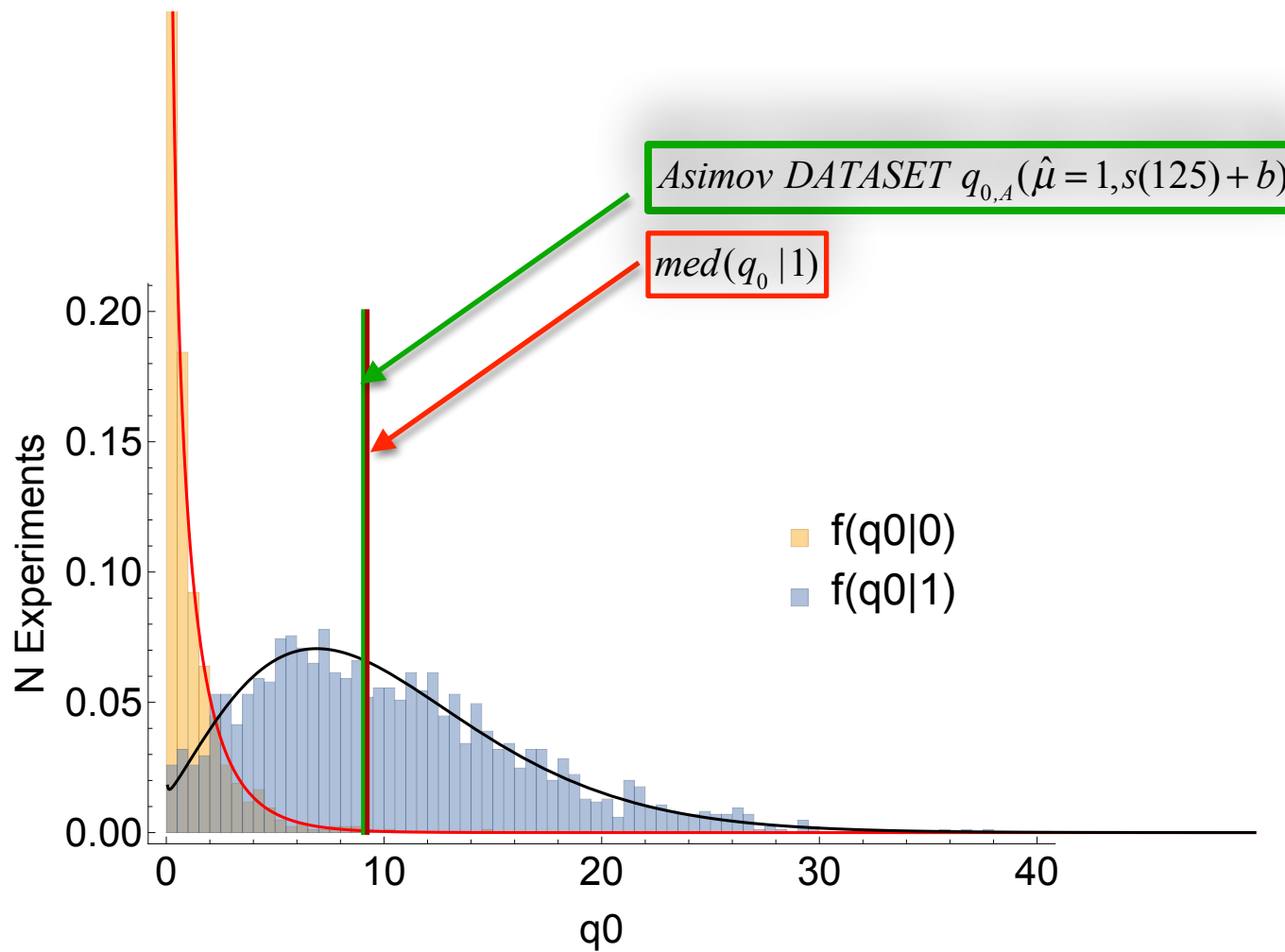
Back to Spin

Distribution of p_0 -value under H_1



$$f(p_0 | H_1) = f([\text{prob}(q \geq q_{obs}) | H_0] | H_1)$$

The Magic of Asimov



q_μ for exclusion

CCGV

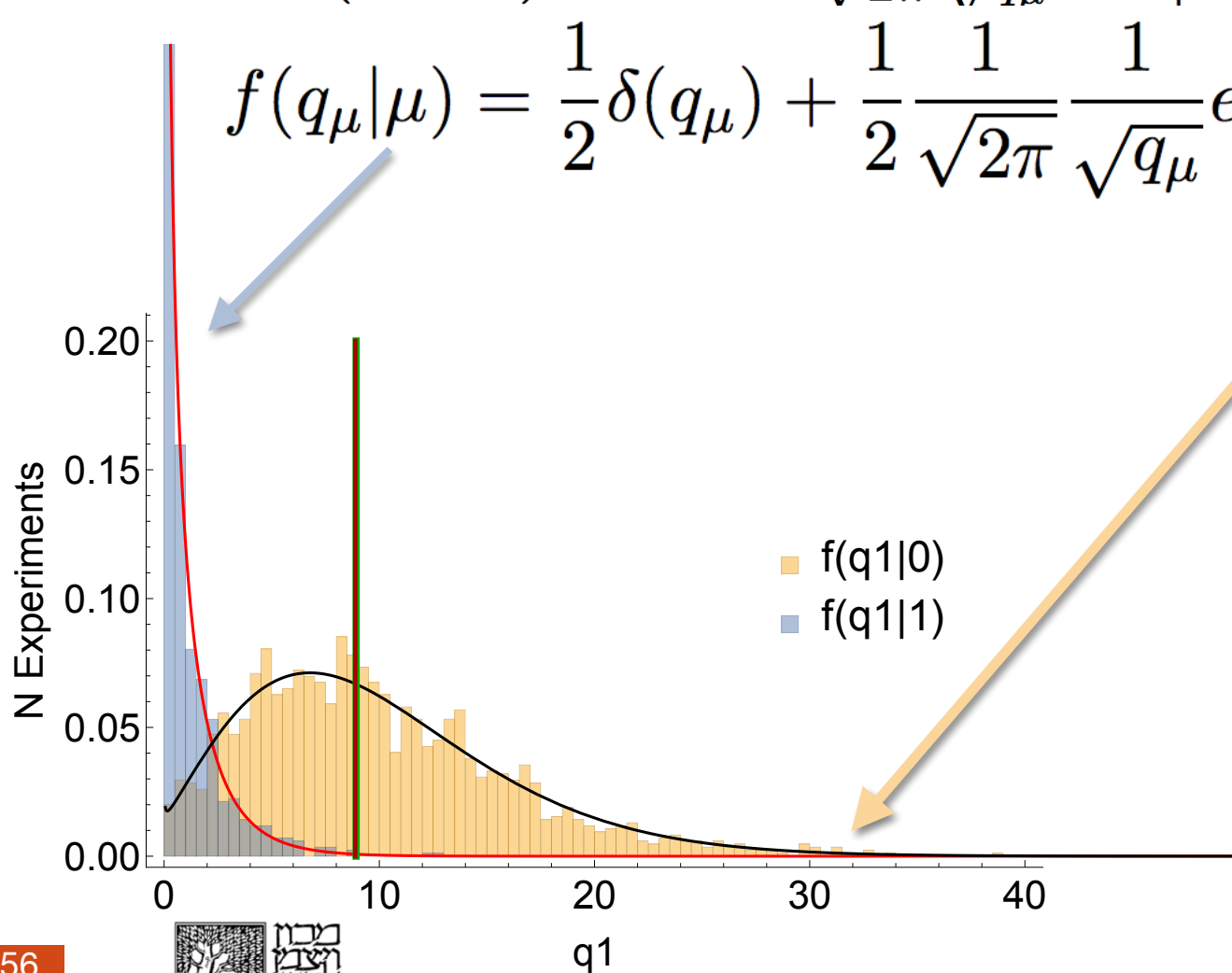
$$q_\mu = \begin{cases} -2 \log \lambda(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases}$$

Upward fluctuations of the signal
do not serve as an evidence against the signal

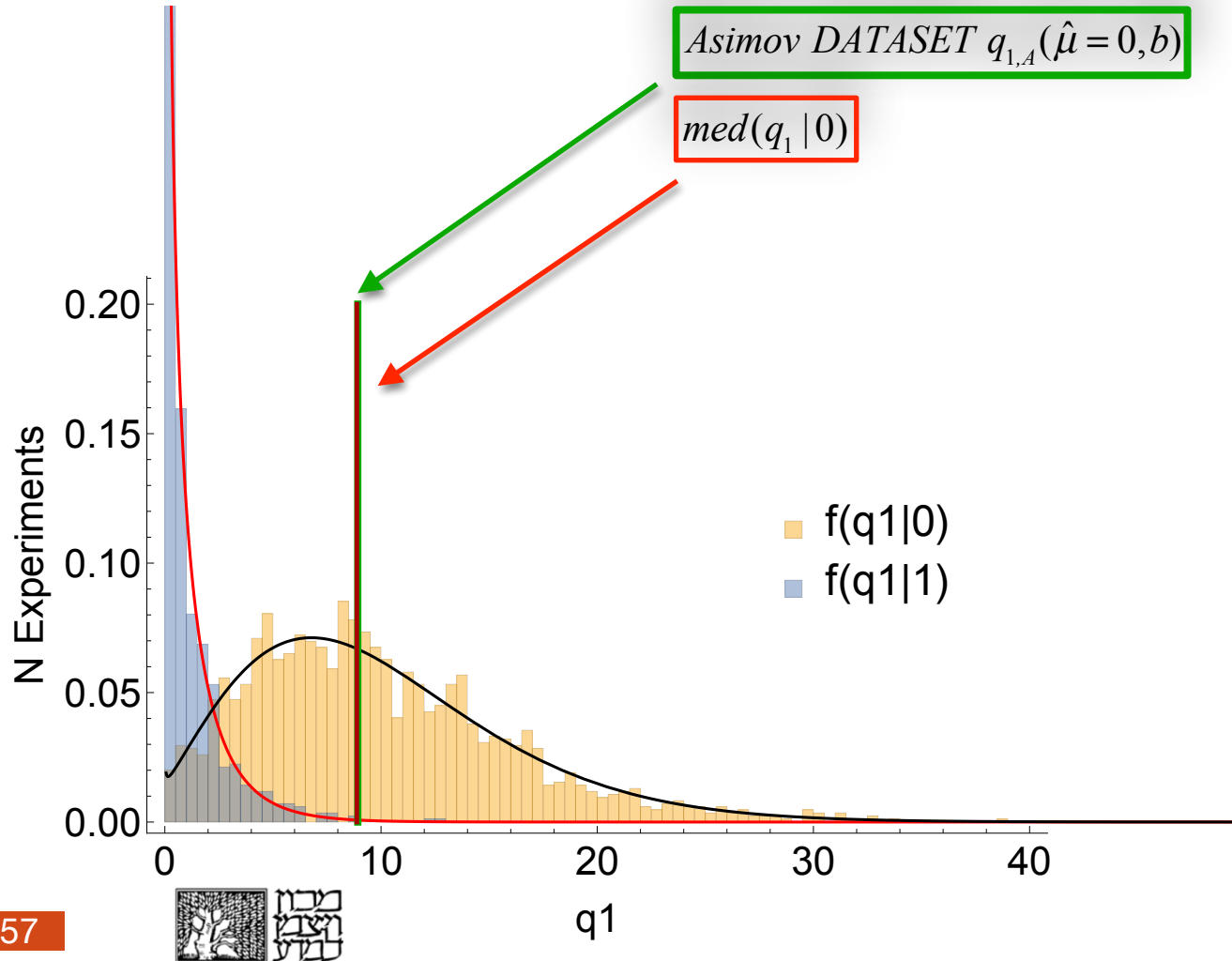
PDF of (q1|1) and (q1|1)

$$f(q_\mu|\mu') = \Phi\left(\frac{\mu' - \mu}{\sigma}\right) \delta(q_\mu) + \frac{1}{2} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{q_\mu}} \exp\left[-\frac{1}{2} \left(\sqrt{q_\mu} - \frac{\mu - \mu'}{\sigma}\right)^2\right]$$

$$f(q_\mu|\mu) = \frac{1}{2} \delta(q_\mu) + \frac{1}{2} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{q_\mu}} e^{-q_\mu/2}$$



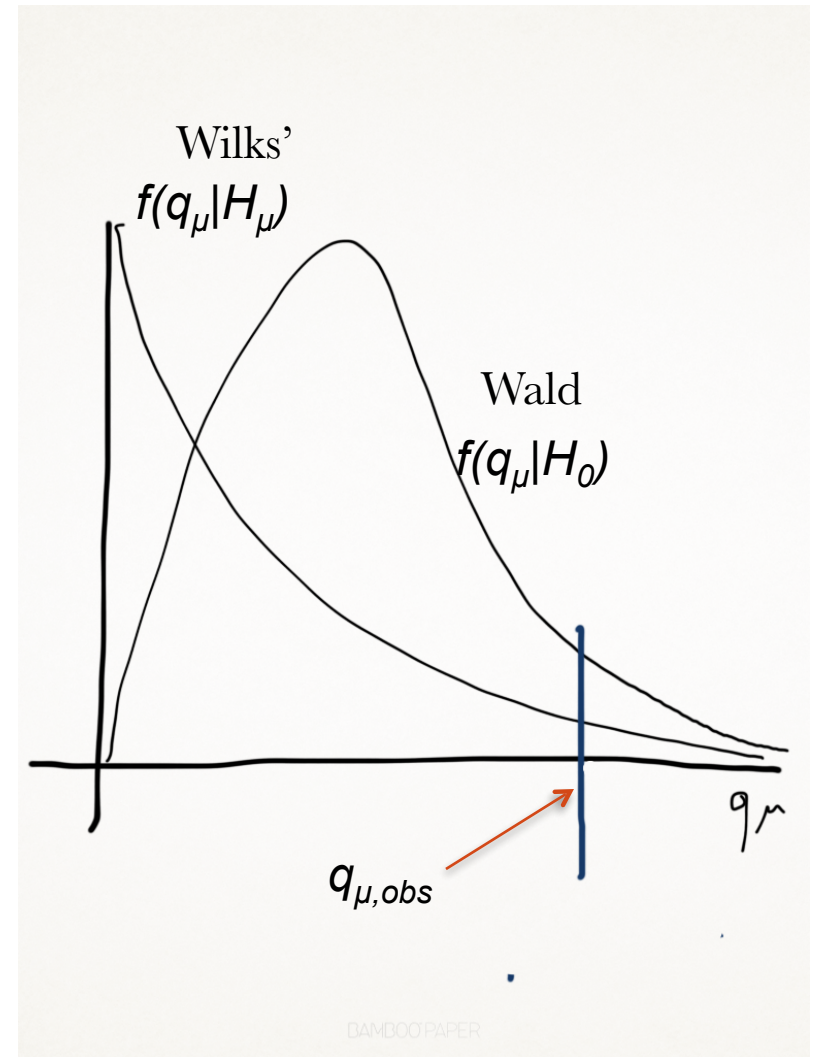
PDF of $(q_1|1)$ and $(q_1|0)$



Exclusion at 95% CL

- We test hypothesis H_μ
- We calculate the PL (profile likelihood) ratio with the one observed data

- $q_{\mu,obs}$

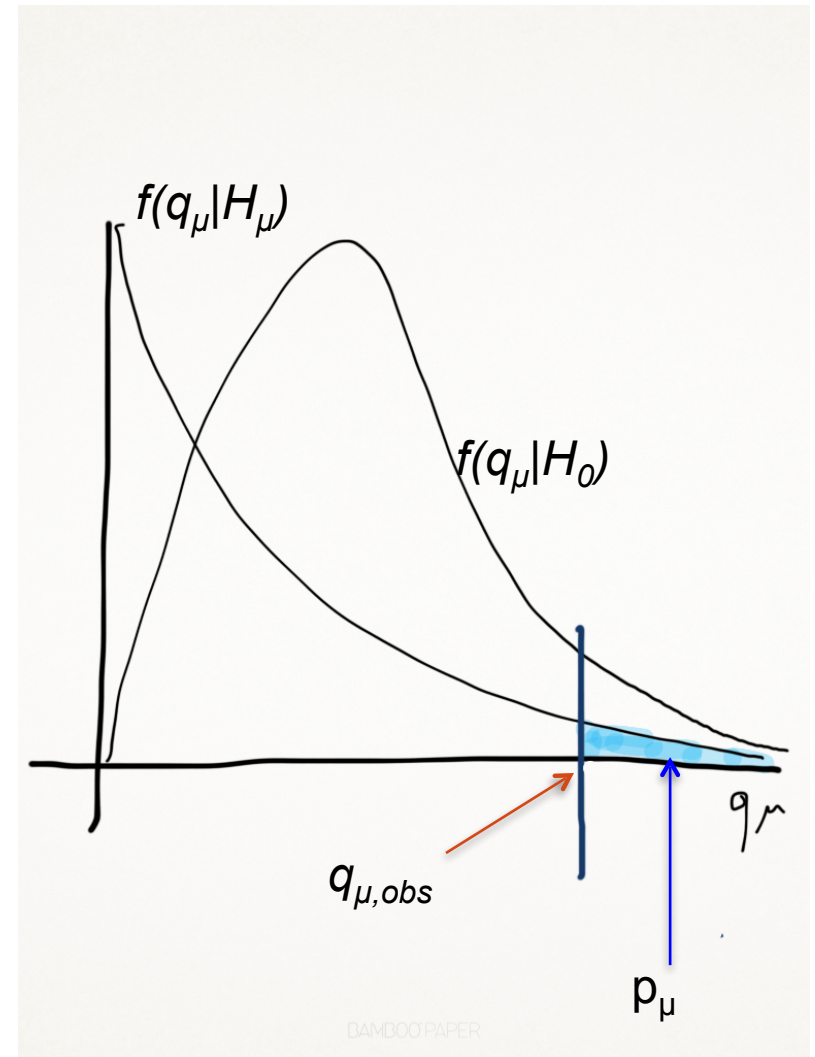


Exclusion at the 95% CL

- Find the p-value of the signal hypothesis H_μ

$$p_\mu = \int_{q_{\mu,obs}}^{\infty} f(q_\mu | H_\mu) dq_\mu$$

- In principle if $p_\mu < 5\%$, H_μ hypothesis is excluded at the 95% CL
- Note that H_μ is for a given Higgs mass m_H



Find μ_{up}

$f(q_\mu | \mu)$

1

0.50

0.10

0.05

0.01

$f(q_\mu | 0)$

$$\text{Let } \langle \hat{\mu} \rangle = 0, \text{ Wald } \rightarrow Z = \sqrt{q_\mu} = \frac{\mu - \hat{\mu}}{\sigma}$$

$$q_{\mu,A} = -2 \ln \frac{L(\mu | 0)}{L(\hat{\mu} = 0 | 0)}$$

$$\sigma_\mu = \frac{\mu}{\sqrt{q_{\mu,A}}}$$

$$p_\mu = 1 - \Phi(\sqrt{q_\mu}) = \alpha \rightarrow \sqrt{q_\mu} = \Phi^{-1}(1 - \alpha)$$

$$\frac{\mu - \hat{\mu}}{\sigma} = \Phi^{-1}(1 - \alpha)$$

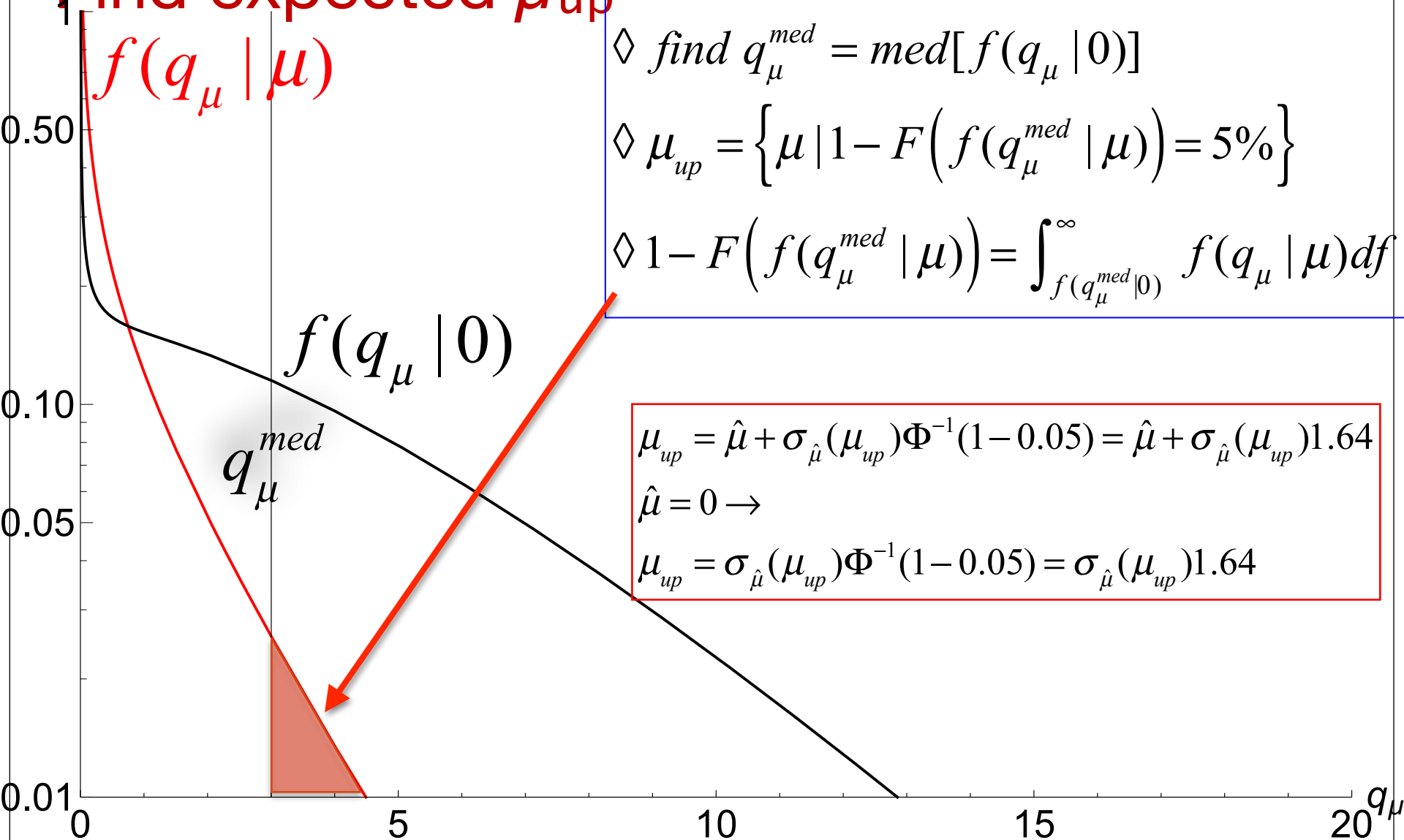
$$\mu_{up} = \{ \mu \mid p_\mu = 5\% \}$$

$$\mu_{up} = \hat{\mu} + \sigma_{\hat{\mu}}(\mu_{up}) \Phi^{-1}(1 - 0.05) = \hat{\mu} + \sigma_{\hat{\mu}}(\mu_{up}) 1.64$$

20 q_μ



Find expected μ_{up}



◇ scan μ

◇ find $q_\mu^{med} = med[f(q_\mu | 0)]$

◇ $\mu_{up} = \left\{ \mu \mid 1 - F\left(f(q_\mu^{med} | \mu)\right) = 5\% \right\}$

◇ $1 - F\left(f(q_\mu^{med} | \mu)\right) = \int_{f(q_\mu^{med} | 0)}^{\infty} f(q_\mu | \mu) df$

$$\mu_{up} = \hat{\mu} + \sigma_{\hat{\mu}}(\mu_{up}) \Phi^{-1}(1 - 0.05) = \hat{\mu} + \sigma_{\hat{\mu}}(\mu_{up}) 1.64$$

$$\hat{\mu} = 0 \rightarrow$$

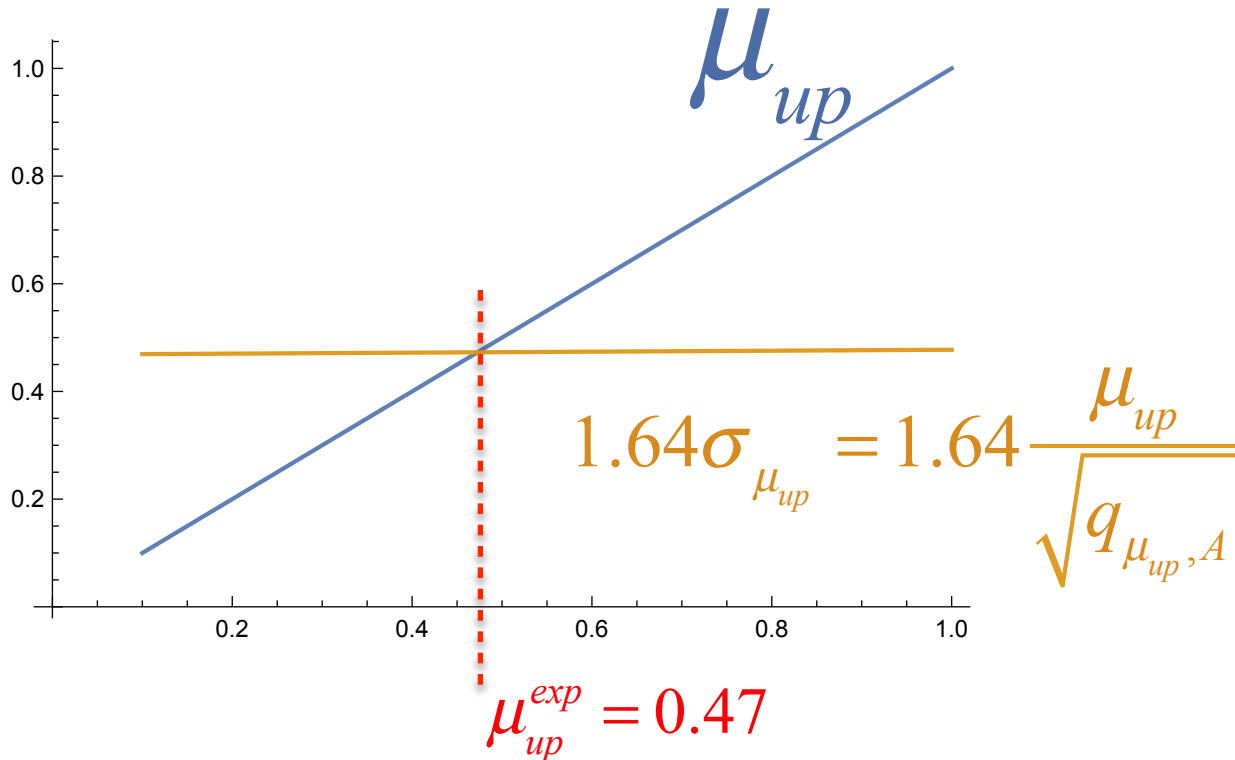
$$\mu_{up} = \sigma_{\hat{\mu}}(\mu_{up}) \Phi^{-1}(1 - 0.05) = \sigma_{\hat{\mu}}(\mu_{up}) 1.64$$

Find expected μ_{up}

$$\mu_{up} = \hat{\mu} + \sigma_{\hat{\mu}}(\mu_{up})\Phi^{-1}(1-0.05) = \hat{\mu} + \sigma_{\hat{\mu}}(\mu_{up})1.64$$

$$\hat{\mu}_A = 0 \rightarrow$$

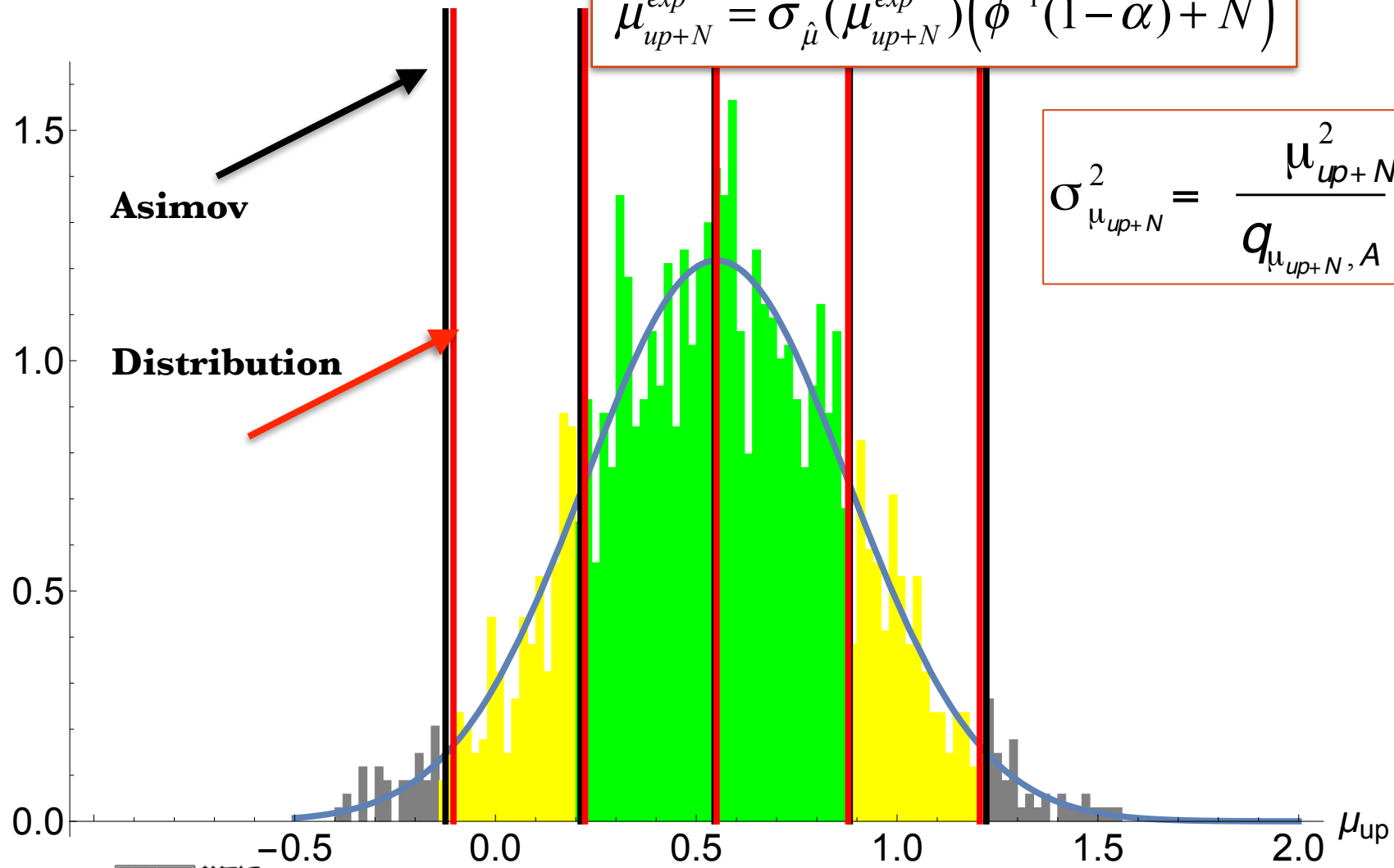
$$\mu_{up}^{exp} = \sigma_{\hat{\mu}}(\mu_{up}^{exp})\Phi^{-1}(1-0.05) = \sigma_{\hat{\mu}}(\mu_{up}^{exp})1.64$$



Expected μ_{up} Bands at m=125

$$\mu_{up+N}^{exp} = \sigma_{\hat{\mu}}(\mu_{up+N}^{exp}) \left(\phi^{-1}(1-\alpha) + N \right)$$

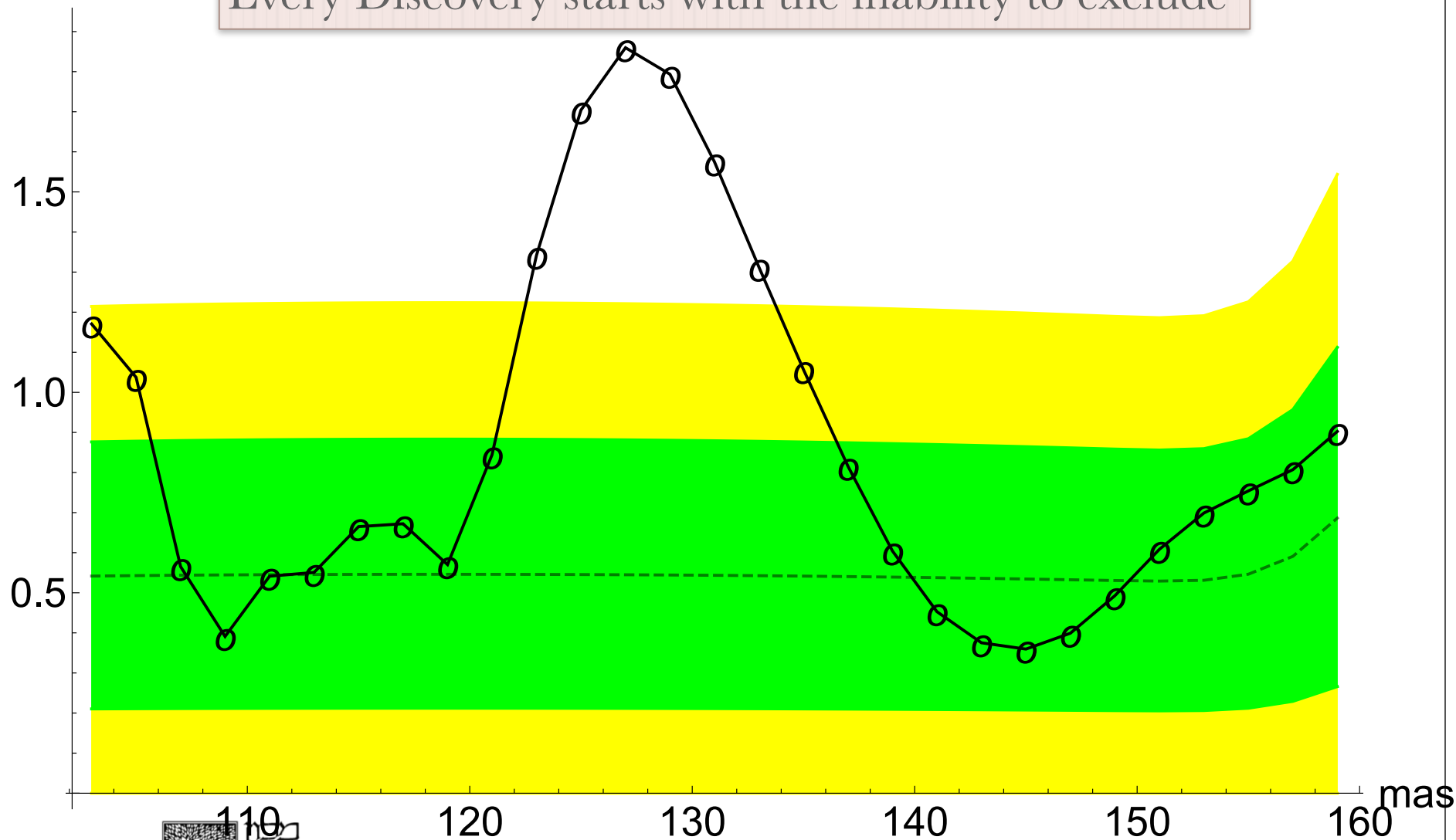
$$\sigma_{\mu_{up+N}}^2 = \frac{\mu_{up+N}^2}{q_{\mu_{up+N}, A}}$$



Brazil Plot

μ_{up}

Every Discovery starts with the inability to exclude



Next: p_0 & Look Elsewhere Effect



E.G., O. Vitells “Trial factors for the look elsewhere effect in high energy physics”,
Eur. Phys. J. C 70 (2010) 525

O. Vitells and E. G., Estimating the significance of a signal in a multi-dimensional search,
1669 Astropart. Phys. 35 (2011) 230, arXiv:1105.4355